

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 0.852781 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-11

Perfect score: 167  
Sequence: 1 LFTFVLDHNECLITGTQSEYTGVAETYSR 31

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	47.3	997	B84726	probable poly(ADP-ribose)
2	53	31.7	726	T21138	hypothetical prote
3	52	31.1	510	S43516	carboxypeptidase C
4	51	30.5	132	B70008	hypothetical prote
5	50.5	30.2	541	G64441	hypothetical prote
6	50	29.9	166	AB0371	conserved hypotet
7	50	29.9	364	T41929	polymerase process
8	50	29.9	384	A84726	probable poly(ADP-ribose)
9	50	29.9	456	A85139	hypothetical prote
10	50	29.9	499	CPBHS	carboxypeptidase C
11	49.5	29.6	1213	A58198	serine/proline-ric
12	49	29.3	152	I47188	Ig heavy chain var
13	48.5	29.0	558	C96933	2-isopropylmalate-2,4
14	48	28.7	250	A69528	capsid-associated
15	48	28.7	474	T10271	NADH2 dehydrogenas
16	48	28.7	521	DNYC27	mRNA capping enzym
17	47.5	28.4	844	QQVZ1	mRNA capping enzym
18	47.5	28.4	844	QOVZVC	probable mRNA guan
19	47.5	28.4	844	T28529	FIR protein - vari
20	47.5	28.4	844	A72162	mRNA capping enzym
21	47.5	28.4	844	T37374	F44E2.1 protein -
22	47.5	28.4	1745	S44816	gag, pol and env p
23	47.5	28.4	2272	T18572	ferric uptake regu
24	47	28.1	150	A42282	hypothetical prote
25	47	28.1	459	AD3223	hypothetical prote
26	47	28.1	488	E95362	hypothetical prote
27	47	28.1	520	AC2416	NADH dehydrogenase
28	47	28.1	670	AI1847	two-component sens
29	47	28.1	885	T26310	hypothetical prote

30	46.5	27.8	164	2	B47207	phycocerythrin alph
31	46.5	27.8	164	2	S73289	phycocerythrin alph
32	46.5	27.8	304	2	AD0864	probable membrane
33	46.5	27.8	401	2	B89863	argininosuccinate
34	46.5	27.8	515	2	A43289	CD44 protein - hu
35	46.5	27.8	796	2	T39962	coatamer complex b
36	46	27.5	77	2	A37317	probable membrane
37	46	27.5	396	2	AC1849	hypothetical prote
38	46	27.5	445	2	T26762	hypothetical prote
39	46	27.5	468	2	AG2899	dihydrolipoamide d
40	46	27.5	468	2	H87674	dihydrolipoamide d
41	46	27.5	491	2	AC1816	hypothetical prote
42	46	27.5	518	2	I39989	spore cortex synth
43	46	27.5	527	2	A86607	lysyl tRNA synthet
44	46	27.5	527	2	E72016	lysine-tRNA ligase
45	46	27.5	601	2	B75136	hypothetical prote

## ALIGNMENTS

### RESULT 1

B84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84726  
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moifatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999  
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84726  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-997 <STO>  
A:Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31870  
A:Map position: 2

Query Match 47.3%; Score 79; DB 2; Length 997;  
Best Local Similarity 48.4%; Pred. NO. 0.0049;  
Matches 15; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY	1	LFTFVLDHNECLITGTQSEYTGVAETYSR	31
DB	274	LFLPRMDDNEATFVGAERFSCYTGYSFR	304

### RESULT 2

T21138  
hypothetical protein F20C5.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T21138  
R:Matthews, P.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19381  
A:Accession: T21138  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-726 <WIL>  
A:Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN000022; CESP:F20C5.1  
A:Experimental source: clone F20C5  
C:Genetics:  
A:Gene: CESP:F20C5.1  
A:Map position: 4  
A:Introns: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1

Query Match 31.7%; Score 53; DB 2; Length 726;

```

Query Match          30.5%; Score 51; DB 2; Length 132;
Best Local Similarity 42.9%; Pred. No. 5,7;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY      8 HNECLIIITGTEQSYEYTGAAE 28
       |||:|||||:|||||:
DB     62 HNVTLILINGELQYOSYSYIQ 82

RESULT 5
GS64441
Hypothetical protein lpg22p homolog - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:CDate: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: G64441
R:R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, A.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Harrison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; PMID:96337999; PMID:8688087
A:Accession: GS6441
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-541 <BUL>
A:Cross-references: GB:U67556; GB:L77117; NID:g1591766; PIDN:AAB99138.1; PID:g1591770; T:
C:Genetics:
A:Map position: FOR1076255-1077880
A:Start codon: TTG
C:Superfamily: hypothetical protein YPL086c

Query Match          30.2%; Score 50.5; DB 2; Length 541;
Best Local Similarity 30.2%; Pred. No. 31;
Matches 13; Conservative 6; Mismatches 9; Indels 15; Gaps 2;

QY      1 LTEFVLDHNE-----CLIIITGTEQY-----SEYTGAAE 28
       :|||:::|||||:|||||:
DB     292 MFKEIFENPDFKPMVKIYPCLIVETGLTYEMMKRGKYKPYRE 334

RESULT 6
ABO371
Conserved hypothetical protein YPO3051 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:CDate: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0371
R:R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB00001; MUID:21470413; PMID:11586360
A:Accession: AB0371
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-166 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92293.1; PID:g15981004; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3051

Query Match          29.9%; Score 50; DB 2; Length 166;
Best Local Similarity 33.3%; Pred. No. 10;
Matches 10; Conservative 5; Mismatches 11; Indels 4; Gaps 1;

QY      2 FTEVLDHNECLIIITGTEQSYEYTGAAETVR 31
       :|||:|||||:|||||:
DB     64 YRAAIENH----LVGTLEQAAYILGLGSTYR 89

RESULT 7
T41929
polymerase processivity factor - human herpesvirus 7 (strain JI)

```



C;Species: human herpesvirus 7  
A;Variety: strain JI  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C;Accession: T41929  
R;Nicholas, J.  
submitted to the EMBL Data Library, December 1995  
A;Description: Determination and analysis of the complete nucleotide sequence of human h  
A;Reference number: Z22022  
A;Accession: T41929  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-364 <NIC>  
A;Cross-references: EMBL:U43400; PIDN:AAC54689.1  
A;Experimental source: strain JI  
C;Genetics:  
A;Note: U27  
C;Superfamily: human herpesvirus 6 P41 protein

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 41.7%; Pred. No. 24;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITTEQSYSEYT 24  
|||:::|||:::|||:::||  
Db 70 LVKLIIHAECYITDHFSTKT 93

RESULT 8  
A84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: A84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varaken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: A84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <STO>  
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AD32286.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31860  
A;Map position: 2

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 36.0%; Pred. No. 24;  
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITTEQSYSEYT 25  
|||:::|||:::|||:::||  
Db 213 LFLPTMEVTRAEIVGGAERFSLYTG 237

RESULT 9  
A85139  
hypothetical protein AT4gl2910 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C;Accession: A85139  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: A85139  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-456 <STO>  
A;Cross-references: GB:NC\_001268; MID:g7267993; PIDN:CAB78333.1; GSPDB:GN00140  
C;Genetics:

A;Gene: AT4gl2910  
A;Map position: 4  
C;Superfamily: serine carboxypeptidase

Query Match 29.9%; Score 50; DB 2; Length 456;  
Best Local Similarity 40.0%; Pred. No. 30;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQSYSEYTYG 26  
|||:::|||:::|||:::||  
Db 391 DHDMCVPTGSSEAWTKSLGY 410

RESULT 10  
CPBHS  
carboxypeptidase C (EC 3.4.16.5) precursor - barley  
N;Alternate names: serine-type carboxypeptidase I  
C;Species: Hordeum vulgare (barley)  
C;Date: 30-Jun-1998 #sequence\_revision 24-Nov-1999 #text\_change 16-Jun-2000  
R;Roche, A.; Lok, F.; Cameron-Millis, V.; von Wettstein, D.  
submitted to the EMBL Data Library, November 1996  
A;Description: The gene family of serine carboxypeptidases in barley.  
A;Reference number: Z15412  
A;Accession: T05367  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-499 <ROC>  
A;Cross-references: EMBL:Y09603; PIDN:CAA70816.1  
A;Experimental source: cv. Himalaya, aleurone  
R;Doan, N.P.; Fincher, G.B.  
J. Biol. Chem. 263, 11106-11110, 1988  
A;Title: The A- and B-chains of carboxypeptidase I from germinated barley originate from  
A;Reference number: A29226; MUID:38298749; PMID:3403516  
A;Accession: A29226  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 88-499 <DOA>  
A;Cross-references: GB:J03897; NID:g167011; PIDN:AAA32940.1; PID:g167012  
R;Sorensen, S.B.; Breddam, K.; Svendsen, I.  
Carlsberg Res. Commun. 51, 475-485, 1986  
A;Title: Primary structure of carboxypeptidase I from malted barley.  
A;Reference number: A30769  
A;Accession: A25858  
A;Molecule type: protein  
A;Residues: 31-101,'P',103-296;352-424,'X',426-499 <SOR>  
C;Genetics:  
A;Gene: Cxpl  
C;Complex: homodimer of two mature products, each containing an A and a B chain  
C;Superfamily: serine carboxypeptidase  
C;Keywords: disulfide bond; glycoprotein; homodimer; hydrolase; serine carboxypeptidase;  
F;1-30/Domain: signal sequence #status predicted <SIG>  
F;31-296,352-499/Product: carboxypeptidase C #status experimental <MAT>  
F;31-296/Domain: chain A #status experimental <CHA>  
F;297-351/Domain: connecting peptide #status predicted <CNP>  
F;352-499/Domain: chain B #status experimental <CHB>  
F;148,262,407/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;188,423,476/Active site: Ser, Asp, His #status predicted

Query Match 29.9%; Score 50; DB 1; Length 499;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQSYSEYTYG 26  
|||:::|||:::|||:::||  
Db 421 DHDMCVPTGSSEAWTKSLGY 440

RESULT 11  
A58198  
serine/proline-rich FEL protein, splice form 1 - human  
N;Alternate names: AF-4 protein  
C;Species: Homo sapiens (man)

C;Species: human herpesvirus 7  
A;Variety: strain JI  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C;Accession: T41929  
R;Nicholas, J.  
submitted to the EMBL Data Library, December 1995  
A;Description: Determination and analysis of the complete nucleotide sequence of human h  
A;Reference number: Z22022  
A;Accession: T41929  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-364 <NIC>  
A;Cross-references: EMBL:U43400; PIDN:AAC54689.1  
A;Experimental source: strain JI  
C;Genetics:  
A;Note: U27  
C;Superfamily: human herpesvirus 6 P41 protein

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 41.7%; Pred. No. 24;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITGTEQSEVYT 24  
||| : : : ||| : : :  
Db 70 LVKLIIHAECYITDHFSTKT 93

RESULT 8  
A84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: A84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varaken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: A84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <STO>  
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AD32286.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31860  
A;Map position: 2

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 36.0%; Pred. No. 24;  
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITGTEQSEVYT 25  
||| : : : ||| : : :  
Db 213 LFLPTMEVTAEIVGAERFSLYTG 237

RESULT 9  
A85139  
hypothetical protein AT4gl2910 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C;Accession: A85139  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: A85139  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-456 <STO>  
A;Cross-references: GB:NC\_001268; MID:g7267993; PIDN:CAB78333.1; GSPDB:GN00140  
C;Genetics:

A;Gene: AT4gl2910  
A;Map position: 4  
C;Superfamily: serine carboxypeptidase

Query Match 29.9%; Score 50; DB 2; Length 456;  
Best Local Similarity 40.0%; Pred. No. 30;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQSYEYTG 26  
||| : : : ||| : : :  
Db 391 DHDMCVPTGSEAWTKSLGY 410

RESULT 10  
CPBHS  
carboxypeptidase C (EC 3.4.16.5) precursor - barley  
N;Alternate names: serine-type carboxypeptidase I  
C;Species: Hordeum vulgare (barley)  
C;Date: 30-Jun-1998 #sequence\_revision 24-Nov-1999 #text\_change 16-Jun-2000  
R;Roche, A.; Lok, F.; Cameron-Mills, V.; von Wettstein, D.  
submitted to the EMBL Data Library, November 1996  
A;Description: The gene family of serine carboxypeptidases in barley.  
A;Reference number: Z15412  
A;Accession: T05367  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-499 <ROC>  
A;Cross-references: EMBL:Y09603; PIDN:CAA70816.1  
A;Experimental source: cv. Himalaya, aleurone  
R;Doan, N.P.; Fincher, G.B.  
J. Biol. Chem. 263, 11106-11110, 1988  
A;Title: The A- and B-chains of carboxypeptidase I from germinated barley originate from  
A;Reference number: A29226; MUID:38298749; PMID:3403516  
A;Accession: A29226  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 88-499 <DOA>  
A;Cross-references: GB:J03897; NID:g167011; PIDN:AAA32940.1; PID:g167012  
R;Sorensen, S.B.; Breddam, K.; Svendsen, I.  
Carlsberg Res. Commun. 51, 475-485, 1986  
A;Title: Primary structure of carboxypeptidase I from malted barley.  
A;Reference number: A30769  
A;Accession: A25858  
A;Molecule type: protein  
A;Residues: 31-101,'P',103-296;352-424,'X',426-499 <SOR>  
C;Genetics:  
A;Gene: Cxpl  
C;Complex: homodimer of two mature products, each containing an A and a B chain  
C;Superfamily: serine carboxypeptidase  
C;Keywords: disulfide bond; glycoprotein; homodimer; hydrolase; serine carboxypeptidase;  
F;1-30/Domain: signal sequence #status predicted <SIG>  
F;31-296,352-499/Product: carboxypeptidase C #status experimental <MAT>  
F;31-296/Domain: chain A #status experimental <CHA>  
F;297-351/Domain: connecting peptide #status predicted <CNP>  
F;352-499/Domain: chain B #status experimental <CHB>  
F;148,262,407/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;188,423,476/Active site: Ser, Asp, His #status predicted

Query Match 29.9%; Score 50; DB 1; Length 499;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQSYEYTG 26  
||| : : : ||| : : :  
Db 421 DHDMCVPTGSEAWTKSLGY 440

RESULT 11  
A58198  
serine/proline-rich FEL protein, splice form 1 - human  
N;Alternate names: AF-4 protein  
C;Species: Homo sapiens (man)

C;Species: human herpesvirus 7  
A;Variety: strain JI  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
C;Accession: T41929  
R;Nicholas, J.  
submitted to the EMBL Data Library, December 1995  
A;Description: Determination and analysis of the complete nucleotide sequence of human h  
A;Reference number: Z22022  
A;Accession: T41929  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-364 <NIC>  
A;Cross-references: EMBL:U43400; PIDN:AAC54689.1  
A;Experimental source: strain JI  
C;Genetics:  
A;Note: U27  
C;Superfamily: human herpesvirus 6 P41 protein

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 41.7%; Pred. No. 24;  
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITTEQSYSEYT 24  
||| : ||| : ||| : ||| :  
Db 70 LVKLIIHAECYITDHFSTKT 93

RESULT 8  
A84726  
probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: A84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varaken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: A84726  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-364 <STO>  
A;Cross-references: GB:AE002093; NID:g4887750; PIDN:AD32286.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g31860  
A;Map position: 2

Query Match 29.9%; Score 50; DB 2; Length 364;  
Best Local Similarity 36.0%; Pred. No. 24;  
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 1 LFTVELDHNECLITTEQSYSEYT 25  
||| : ||| : ||| : ||| :  
Db 213 LFLPTMEVTRAEIVGGAERFSLYTG 237

RESULT 9  
A85139  
hypothetical protein AT4gl2910 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C;Accession: A85139  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: A85139  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-456 <STO>  
A;Cross-references: GB:NC\_001268; MID:g7267993; PIDN:CAB78333.1; GSPDB:GN00140  
C;Genetics:

A;Gene: AT4gl2910  
A;Map position: 4  
C;Superfamily: serine carboxypeptidase

Query Match 29.9%; Score 50; DB 2; Length 456;  
Best Local Similarity 40.0%; Pred. No. 30;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQYSYTYG 26  
||| : ||| : ||| : ||| :  
Db 391 DHDMCVPTGSSEAWTKSLGY 410

RESULT 10  
CPBHS  
carboxypeptidase C (EC 3.4.16.5) precursor - barley  
N;Alternate names: serine-type carboxypeptidase I  
C;Species: Hordeum vulgare (barley)  
C;Date: 30-Jun-1998 #sequence\_revision 24-Nov-1999 #text\_change 16-Jun-2000  
R;Roche, A.; Lok, F.; Cameron-Millis, V.; von Wettstein, D.  
submitted to the EMBL Data Library, November 1996  
A;Description: The gene family of serine carboxypeptidases in barley.  
A;Reference number: Z15412  
A;Accession: T05367  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-499 <ROC>  
A;Cross-references: EMBL:Y09603; PIDN:CAA70816.1  
A;Experimental source: cv. Himalaya, aleurone  
R;Doan, N.P.; Fincher, G.B.  
J. Biol. Chem. 263, 11106-11110, 1988  
A;Title: The A- and B-chains of carboxypeptidase I from germinated barley originate from  
A;Reference number: A29226; MUID:38298749; PMID:3403516  
A;Accession: A29226  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 88-499 <DOA>  
A;Cross-references: GB:J03897; NID:g167011; PIDN:AAA32940.1; PID:g167012  
R;Sorensen, S.B.; Breddam, K.; Svendsen, I.  
Carlsberg Res. Commun. 51, 475-485, 1986  
A;Title: Primary structure of carboxypeptidase I from malted barley.  
A;Reference number: A90769  
A;Accession: A25858  
A;Molecule type: protein  
A;Residues: 31-101,'P',103-296;352-424,'X',426-499 <SOR>  
C;Genetics:  
A;Gene: Cxpl  
C;Complex: homodimer of two mature products, each containing an A and a B chain  
C;Superfamily: serine carboxypeptidase  
C;Keywords: disulfide bond; glycoprotein; homodimer; hydrolase; serine carboxypeptidase;  
F;1-30/Domain: signal sequence #status predicted <SIG>  
F;31-296,352-499/Product: carboxypeptidase C #status experimental <MAT>  
F;31-296/Domain: chain A #status experimental <CHA>  
F;297-351/Domain: connecting peptide #status predicted <CNP>  
F;352-499/Domain: chain B #status experimental <CHB>  
F;148,262,407/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;188,423,476/Active site: Ser, Asp, His #status predicted

Query Match 29.9%; Score 50; DB 1; Length 499;  
Best Local Similarity 40.0%; Pred. No. 33;  
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 7 DHNECLIITGEQYSYTYG 26  
||| : ||| : ||| : ||| :  
Db 421 DHDMCVPTGSSEAWTKSLGY 440

RESULT 11  
A58198  
serine/proline-rich FEL protein, splice form 1 - human  
N;Alternate names: AF-4 protein  
C;Species: Homo sapiens (man)

C;Date: 02-Jul-1996 #sequence\_revision 25-Oct-1996 #text\_change 01-Dec-2000  
C;Accession: A58198; I52572  
R;Morrissey, J. J. M.  
Blood 81, 1124-1131, 1993  
A;Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.  
A;Reference number: I52572; MUID:93184301; PMID:8443374  
A;Accession: A58198  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-1213 <MOR>  
A;Cross-references: GB:L25050; NID:g407323  
A;Accession: I52572  
A;Status: translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 'MAAQS', 13-1213 <M02>  
A;Cross-references: GB:L25050; NID:g407323; PIDN:AAA36642.1; PID:g407324  
C;Comment: This protein is one of several to form chimeric fusion proteins with the trit  
C;Genetics:  
A;Gene: GDB:MLT2; AF-4; FEL  
A;Cross-references: GDB:I36792; OMIM:159557  
A;Map position: 4q21-4q21  
C;Keywords: alternative splicing

Query Match 29.6%; Score 49.5; DB 2; Length 1213;  
Best Local Similarity 46.4%; Pred. No. 1e+02;  
Matches 13; Conservative 3; Mismatches 9; Indels 3; Gaps 1;

QY 5 VLDHNECLITGTQYSB---VTGYAET 29  
Db 1002 VLSFNEGIATESQSQSKSAISYVSET 1029

RESULT 12  
I47188  
Ig heavy chain variable VDJ region - pig (fragment)  
C;Species: Sus scrofa domestica (domestic pig)  
C;Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
C;Accession: I47188  
R;Sun, J.; Kacskovics, I.; Brown, W.R.; Butler, J.E.  
J. Immunol. 153, 5618-5627, 1994  
A;Title: Expressed swine VH genes belong to a small VH gene family homologous to human V  
A;Reference number: I47177; MUID:95081609; PMID:7989761  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-152 <SUN>  
A;Cross-references: EMBL:U15447; NID:g571380; PIDN:AAA67013.1; PID:g571381  
C;Superfamily: immunoglobulin V region; immunoglobulin homology  
F:34-119/Domain: immunoglobulin homology <IMM>

Query Match 29.3%; Score 49; DB 2; Length 152;  
Best Local Similarity 40.9%; Pred. No. 13;  
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 10 ECLITGTQYSYVGYAET 31  
Db 65 ECLASIGSSSYSGHTYYADSVK 86

RESULT 13  
C96933  
2-isopropylmalate synthase [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C;Accession: C96933  
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: C96933  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-558 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK78254.1; PID:g15023111; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0273  
C;Superfamily: 2-isopropylmalate synthase

Query Match 29.0%; Score 48.5; DB 2; Length 558;  
Best Local Similarity 37.5%; Pred. No. 61;  
Matches 12; Conservative 7; Mismatches 10; Indels 3; Gaps 1;

QY 1 LFTVELDHNECLITGTPE---QYSEYTYGAET 29  
Db 136 VFKDKDEKSLAKGAQVMKKYSEYTYSES 167

RESULT 14  
A69528  
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (hpcB-2) homolog - Archaeoglobus fulgidus  
C;Species: Archaeoglobus fulgidus  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 29-Sep-1999  
C;Accession: A69528  
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
A;Reference number: A69250; MUID:98049343; PMID:9389475  
A;Accession: A69528  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-250 <KLE>  
A;Cross-references: GB:AF000952; GB:AE000782; NID:g2689275; PIDN:AA889031.1; PID:g264830  
C;Superfamily: 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase

Query Match 28.7%; Score 48; DB 2; Length 250;  
Best Local Similarity 34.8%; Pred. No. 31;  
Matches 8; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 3 TEVLHDHNECLITGTQYSEYTG 25  
Db 76 TAVIGHDDCILPQISQVDYEG 98

RESULT 15  
TI0271  
capsid-associated protein 1629 - Orgyia pseudotsugata nuclear polyhedrosis virus  
C;Species: Orgyia pseudotsugata nuclear polyhedrosis virus, OpMNPV  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 11-May-2000  
C;Accession: TI0271  
R;Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohrmann, G.F.  
Virology 229, 381-399, 1997  
A;Title: The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis  
A;Reference number: 217011; MUID:97271300; PMID:9126251  
A;Accession: TI0271  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-474 <AHR>  
A;Cross-references: EMBL:U75930; NID:g2934903; PID:g2934904

Query Match 28.7%; Score 48; DB 2; Length 474;  
Best Local Similarity 31.2%; Pred. No. 60;  
Matches 10; Conservative 10; Mismatches 10; Indels 2; Gaps 1;

QY 1 LFTVELDHNECLITGT---EQYSEYTYGAET 30  
Db 116 LRVDVNDAAEAVNLSDGFYEYSKVISYQQTFF 147

Search completed: May 26, 2004, 18:47:57  
Job time : 1.85278 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 111.936 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGGCGPCTKPRWDAAA.....YHVESCTQTNNQPGQRTGA 977

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5184	100.0	977	AAE25629	Bovine po
2	5184	100.0	977	Aau76020	Bovine po
3	5184	100.0	977	Aau75799	Bovine po
4	5184	100.0	977	Abg72278	Bovine po
5	4635.5	89.4	976	AAE25630	Human pol
6	4635.5	89.4	976	Aau76021	Human pol
7	4635.5	89.4	976	Aau76012	Human pol
8	4635.5	89.4	976	Abg72279	Human pol
9	4332.5	83.6	968	AAE25631	Murine po
10	4332.5	83.6	968	Aau76022	Mouse pol
11	4332.5	83.6	968	AAU76013	Mouse pol
12	4332.5	83.6	968	Abg72280	Murine p
13	1063.5	20.5	768	AAE25632	Fruit fly
14	1063.5	20.5	768	AAE25632	Fruit fly
15	1063.5	20.5	768	AAU76023	Fruit fly
16	1063.5	20.5	768	AAU76014	Fruit fly
17	1063.5	20.5	768	Abg72281	Fruit fly
18	465	9.0	726	AAE25633	Poly aden
19	465	9.0	726	AAU76024	Worm poly
20	465	9.0	726	AAU76015	Worm poly
21	465	9.0	726	Abg72282	C. elegans
22	335.5	6.5	819	ABG20721	Novel hum
23	201	3.9	100	ABG11103	Novel hum
24	178	3.4	33	AAE25652	Bovine po
25	174.5	3.4	2194	AAU40114	Human pol

26	174	3.4	1046	4	ABG10125	Novel hum
27	170	3.3	1284	4	ABG09663	Novel hum
28	170	3.3	1284	4	ABG18482	Novel hum
29	170	3.3	1284	4	ABG06083	Novel hum
30	170	3.3	1284	4	ABG10823	Novel hum
31	170	3.3	1300	4	ABG09142	Novel hum
32	170	3.3	1308	4	ABG04003	Novel hum
33	167	3.2	31	5	AAE25634	Bovine po
34	167	3.2	31	5	AAE25651	Bovine po
35	167	3.2	31	5	AAU76025	Bovine po
36	167	3.2	31	5	AAU76016	Bovine po
37	167	3.2	31	6	ABG72283	Oligopept
38	166	3.2	962	4	ABG04842	Novel hum
39	166	3.2	1462	6	ABP58346	Human cel
40	166	3.2	1462	6	ABU05132	Human exp
41	166	3.2	1462	6	ABU05136	Human exp
42	166	3.2	1462	6	ABU05131	Human exp
43	166	3.2	1462	6	ABU05135	Human exp
44	166	3.2	1524	4	ABG04845	Novel hum
45	164.5	3.2	1278	4	ABB71882	Drosophil

ALIGNMENTS

RESULT 1

AAE25629  
ID AAE25629 standard; protein; 977 AA.

XX AAE25629;

AC AAE25629;

DT 04-NOV-2002 (first entry)

XX Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).

XX Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nontropic;  
KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
XX antisenase therapy.

XX Bos taurus.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson BL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAU42081.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 3; Col 47-45; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG  
 XX  
 SQ Sequence 977 AA;

Query Match 100.0%; Score 5184; DB 5; Length 977;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAGPGCEPCTKRPRWDAATSPPAASDARSFPQQRRLVDSKADPQVFRVPPSSGAL 60  
 DB 1 MSAGPGCEPCTKRPRWDAATSPPAASDARSFPQQRRLVDSKADPQVFRVPPSSGAL 60

QY 61 GRAGQHRGSATSLVFKQKTIISWMDTKIKTVESLHSEKNNNTREESMSSVQKDNFY 120  
 DB 61 GRAGQHRGSATSLVFKQKTIISWMDTKIKTVESLHSEKNNNTREESMSSVQKDNFY 120

QY 121 QHNNEKLENSQLGFDKSPVEKGTQYLKQHQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
 DB 121 QHNNEKLENSQLGFDKSPVEKGTQYLKQHQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180

QY 181 EQFSNANVDQSPKDDHSDTNSERSDNOQPLTHVLANAKQTWDEQGRARSHQCKG 240  
 DB 181 EQFSNANVDQSPKDDHSDTNSERSDNOQPLTHVLANAKQTWDEQGRARSHQCKG 240

QY 241 ACHPAECAGCQOQRETVVSSPLSDTGSDEVGTGLKNNRLNQESSLGNSPPFKESE 300  
 DB 241 ACHPAECAGCQOQRETVVSSPLSDTGSDEVGTGLKNNRLNQESSLGNSPPFKESE 300

QY 301 PESPMVDVNSKNSQDSSEADSETSPGFDEQEDSSSAQTANKPSFPQPREADTELKRS 360  
 DB 301 PESPMVDVNSKNSQDSSEADSETSPGFDEQEDSSSAQTANKPSFPQPREADTELKRS 360

QY 361 KGGEIRLHFQEGGSRAGMDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420  
 DB 361 KGGEIRLHFQEGGSRAGMDVNAKRPSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420

QY 421 AEDKEKECEMKHOKTERKIPKIPPHLSPDKMLGTPIEMRRMPRCGIRLPLRPSAN 480  
 DB 421 AEDKEKECEMKHOKTERKIPKIPPHLSPDKMLGTPIEMRRMPRCGIRLPLRPSAN 480

QY 481 HTVTIRVLLRIGEVKPFPTTHFKDLWKNKXMPCEQNLYPVEDENGERRAAGSRWELI 540  
 DB 481 HTVTIRVLLRIGEVKPFPTTHFKDLWKNKXMPCEQNLYPVEDENGERRAAGSRWELI 540

QY 541 QTALLNRLTRPONLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEAOHLQSLIPDMVK 600  
 DB 541 QTALLNRLTRPONLKDAILKYNVAYSKKWDFTALIDFWDKVLEAEAOHLQSLIPDMVK 600

QY 601 IALCLPNICTOPIPLLKQNNHSITMSQEQIASILANAFCTFFPRNAKMSSEYSSYDI 660  
 DB 601 IALCLPNICTOPIPLLKQNNHSITMSQEQIASILANAFCTFFPRNAKMSSEYSSYDI 660

QY 661 NFNLFEGRSRKPKKLTLCFYFRFVTEKKTGLVTFTRQSLDFPFEWERCEKLLTRLH 720  
 DB 661 NFNLFEGRSRKPKKLTLCFYFRFVTEKKTGLVTFTRQSLDFPFEWERCEKLLTRLH 720

QY 721 VTYEGTIEGNGQMLQVDFANRFGVGVTSAGLVOEETREFLINPELIVSRLPTEVLDRNE 780  
 DB 721 VTYEGTIEGNGQMLQVDFANRFGVGVTSAGLVOEETREFLINPELIVSRLPTEVLDRNE 780

QY 781 CLIIITGTEQYSYVTGVAETRYRWARSHEDRSERDDWQRTTEIVAIDALHFRYLDQFVPE 840  
 DB 781 CLIIITGTEQYSYVTGVAETRYRWARSHEDRSERDDWQRTTEIVAIDALHFRYLDQFVPE 840

QY 841 KIRRELKAYCGFLRPGVSENLSAVATGNGWCGAFGGDARLKALIQILAAVAERDVVY 900  
 DB 841 KIRRELKAYCGFLRPGVSENLSAVATGNGWCGAFGGDARLKALIQILAAVAERDVVY 900

QY 901 FTFGDSELMEDIYSMTFLTERKLTGVEVYKLLRYNNEECRCNSTPGDILKLYPFIYHA 960  
 DB 901 FTFGDSELMEDIYSMTFLTERKLTGVEVYKLLRYNNEECRCNSTPGDILKLYPFIYHA 960

QY 961 VESCTQTNTNQPGQRTGA 977  
 DB 961 VESCTQTNTNQPGQRTGA 977

RESULT 2  
 AAU76020  
 ID AAU76020 standard; protein; 977 AA.  
 XX  
 AC AAU76020;  
 XX  
 DT  
 XX 08-MAY-2002 (first entry)  
 XX Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 601..617  
 FT /note= "Represents PARG oligopeptide #1"  
 FT Region 761..770  
 FT /note= "Represents PARG oligopeptide #2"  
 FT Region 771..801  
 FT /note= "Represents PARG oligopeptide #3"  
 FT Region 849..880  
 FT /note= "Represents PARG oligopeptide #4"  
 XX  
 XX US6337202-B1.  
 XX  
 XX 08-JAN-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511477.  
 XX  
 XX 01-MAY-1998; 98US-0093768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.  
 XX N-PSDB; ABK14931.  
 XX  
 XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 XX which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 XX treating neoplastic and neurological disorders, heart attack and stroke.  
 XX  
 XX Claim 2; Col 47-52; 81pp; English.  
 XX  
 XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 XX (PARG) protein which catalyses release of ADP-ribose from an ADP  
 XX (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 XX invention is useful for generating antibodies and can be inhibited or  
 XX activated for diagnosing and treating neoplastic disorders such as  
 XX adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma.  
 XX teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 XX ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 XX disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 XX and related conditions. PARG levels may be enhanced to suppress DNA  
 XX repair and increase the cell's susceptibility to chemotherapy drugs.  
 XX Antagonists of PARG are administered to treat or prevent neoplastic



Matches	977;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MSAGPGCEPCTKRPRWDAATSPPAASDARGPQRRVLDSDAPVQFRVPPSSGGAL	60						
Dd	1	MSAGPGCEPCTKRPRWDAATSPPAASDARGPQRRVLDSDAPVQFRVPPSSGGAL	60						
Qy	61	GRAGQHRGSATSLVFKQKTIISMDTKIGIKTVESELSHKNNNTREESMMSSVQKDNFY	120						
Dd	61	GRAGQHRGSATSLVFKQKTIISMDTKIGIKTVESELSHKNNNTREESMMSSVQKDNFY	120						
Qy	121	QHNNEKLENSQLGFDKSPVEKCTGYLKHOTAAWKWQNGPSEHLESEPPAVTLVP	180						
Dd	121	QHNNEKLENSQLGFDKSPVEKCTGYLKHOTAAWKWQNGPSEHLESEPPAVTLVP	180						
Qy	181	EOFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKOTWEDEGRARSHQKCGK	240						
Dd	181	EOFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKOTWEDEGRARSHQKCGK	240						
Qy	241	ACHPAEACAGCOQEBETDVVSEPLSDTSGEDVGTGLKNNRLNQESSLGNSPPEKESE	300						
Dd	241	ACHPAEACAGCOQEBETDVVSEPLSDTSGEDVGTGLKNNRLNQESSLGNSPPEKESE	300						
Qy	301	PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRPOPREADTELKRGSSA	360						
Dd	301	PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRPOPREADTELKRGSSA	360						
Qy	361	KGGEIRLHFQEGGESRAGMDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK	420						
Dd	361	KGGEIRLHFQEGGESRAGMDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK	420						
Qy	421	AEDKRKEQCEMKHQRTERKIPKIPPHLSPOKMLGTPIEMWRMPRCGIRLPLRPSAN	480						
Dd	421	AEDKRKEQCEMKHQRTERKIPKIPPHLSPOKMLGTPIEMWRMPRCGIRLPLRPSAN	480						
Qy	481	HTVTIRVOLLRIGEVPEKPFTHFDKLDWKNKVMPCSQNLVPVEDENGERAAGRWELI	540						
Dd	481	HTVTIRVOLLRIGEVPEKPFTHFDKLDWKNKVMPCSQNLVPVEDENGERAAGRWELI	540						
Qy	541	QTALENLRLTRPNLKDAILKNNVAYSKKWDFTALIDFWDKVLSEAEAOHLQOSILPDMVK	600						
Dd	541	QTALENLRLTRPNLKDAILKNNVAYSKKWDFTALIDFWDKVLSEAEAOHLQOSILPDMVK	600						
Qy	601	IACLPNICTQPIPLLKOKNHISITMSQEQIASLLANAFCTFPRNNAKMSSESYSDI	660						
Dd	601	IACLPNICTQPIPLLKOKNHISITMSQEQIASLLANAFCTFPRNNAKMSSESYSDI	660						
Qy	661	NFNRLPEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH	720						
Dd	661	NFNRLPEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH	720						
Qy	721	VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLPFTEVLHNE	780						
Dd	721	VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLPFTEVLHNE	780						
Qy	781	CLIIITGTEQSYTGAETRYRWARSHEDSRDDQRRTTEIVAIDALHFRYLDQFVPE	840						
Dd	781	CLIIITGTEQSYTGAETRYRWARSHEDSRDDQRRTTEIVAIDALHFRYLDQFVPE	840						
Qy	841	KIRRELNKAYCGFLRPGVSSENLAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY	900						
Dd	841	KIRRELNKAYCGFLRPGVSSENLAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY	900						
Qy	901	FTFGDSELMRDIYSMHTFLTERKLTGVGVYKLLRLYYNEECRNCSTPGPDIKLYPIYHA	960						
Dd	901	FTFGDSELMRDIYSMHTFLTERKLTGVGVYKLLRLYYNEECRNCSTPGPDIKLYPIYHA	960						
Qy	961	VESCTQTTNQPGQRTGA	977						
Dd	961	VESCTQTTNQPGQRTGA	977						

ID	ABG72278 standard; protein; 977 AA.
XX	ABG72278;
AC	13-MAR-2003 (first entry)
DT	Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.
DE	Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression; cellular response; DNA damage; neoplastic disorder inducing agent; oxidative stress; neoplastic disorder; myocardial infarction; vascular stroke; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; inborn genetic error; reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytosolic; neuroprotective; nootropic; antiparkinsonian; cardiant; vasotropic; anticonvulsant; cerebroprotective; enzyme.
OS	Bos taurus.
XX	US2002132328-A1.
FN	19-SEP-2002.
XX	09-OCT-2001; 2001US-00973451.
PF	01-MAY-1998; 98US-0083768P.
XX	30-APR-1999; 99US-00302812.
PR	(JACO/) JACOBSON M K.
PA	(JACO/) JACOBSON E L.
PA	(AMEJ/) AME J.
PA	(LINW/) LIN W.
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;
PI	WPI; 2003-155895/15.
XX	N-PSDB; ABX14477.
DR	New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)
PT	glycohydrolase activity, for preventing, treating, or ameliorating a
PT	disease condition, e.g. neoplastic disorder, myocardial infarction or
PT	vascular stroke.
XX	Claim 28; Fig 16; 86pp; English.
PS	The present invention relates to the isolation of poly(ADP-ribose)
CC	glycohydrolase (PARG) from several species, and the polynucleotide
CC	sequences encoding them. Methods for inhibiting PARG expression or
CC	overexpressing PARG are also disclosed. PARG is involved in the cellular
CC	response to DNA damage, and is associated with the body's response to
CC	neoplastic disorder inducing agents and oxidative stress. The
CC	polynucleotide sequences encoding PARG and PARG modulators are useful for
CC	preventing, treating, or ameliorating diseases such as neoplastic
CC	disorders, myocardial infarction, vascular stroke, neurodegenerative
CC	diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC	disease), inborn genetic errors, reperfusion following ischaemia, aging,
CC	and neurotoxicity. The polynucleotide sequences are also useful in gene
CC	therapy. The methods are useful for identifying an agent that modulates
CC	PARG activity, identifying a mutant PARG allele in an individual, or
CC	screening candidate molecules for PARG modulating activity. The present
CC	sequence represents bovine PARG enzyme
XX	Sequence 977 AA;
SQ	Query Match 100.0%; Score 5184; DB 6; Length 977;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 977; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 MSAGPGCEPCTKRPRWDAATSPPAASDARGPQRRVLDSDAPVQFRVPPSSGGAL 60
Dd	1 MSAGPGCEPCTKRPRWDAATSPPAASDARGPQRRVLDSDAPVQFRVPPSSGGAL 60
Qy	61 GRAGQHRGSATSLVFKQKTIISMDTKIGIKTVESELSHKNNNTREESMMSSVQKDNFY 120
Dd	61 GRAGQHRGSATSLVFKQKTIISMDTKIGIKTVESELSHKNNNTREESMMSSVQKDNFY 120
Qy	121 QHNNEKLENSQLGFDKSPVEKCTGYLKHOTAAWKWQNGPSEHLESEPPAVTLVP 180
Dd	121 QHNNEKLENSQLGFDKSPVEKCTGYLKHOTAAWKWQNGPSEHLESEPPAVTLVP 180
Qy	181 EOFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKOTWEDEGRARSHQKCGK 240
Dd	181 EOFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKOTWEDEGRARSHQKCGK 240
Qy	241 ACHPAEACAGCOQEBETDVVSEPLSDTSGEDVGTGLKNNRLNQESSLGNSPPEKESE 300
Dd	241 ACHPAEACAGCOQEBETDVVSEPLSDTSGEDVGTGLKNNRLNQESSLGNSPPEKESE 300
Qy	301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRPOPREADTELKRGSSA 360
Dd	301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRPOPREADTELKRGSSA 360
Qy	361 KGGEIRLHFQEGGESRAGMDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420
Dd	361 KGGEIRLHFQEGGESRAGMDVNAKRPGSTSSLNVECRNSKHGRKDSKITDHFMRVPK 420
Qy	421 AEDKRKEQCEMKHQRTERKIPKIPPHLSPOKMLGTPIEMWRMPRCGIRLPLRPSAN 480
Dd	421 AEDKRKEQCEMKHQRTERKIPKIPPHLSPOKMLGTPIEMWRMPRCGIRLPLRPSAN 480
Qy	481 HTVTIRVOLLRIGEVPEKPFTHFDKLDWKNKVMPCSQNLVPVEDENGERAAGRWELI 540
Dd	481 HTVTIRVOLLRIGEVPEKPFTHFDKLDWKNKVMPCSQNLVPVEDENGERAAGRWELI 540
Qy	541 QTALENLRLTRPNLKDAILKNNVAYSKKWDFTALIDFWDKVLSEAEAOHLQOSILPDMVK 600
Dd	541 QTALENLRLTRPNLKDAILKNNVAYSKKWDFTALIDFWDKVLSEAEAOHLQOSILPDMVK 600
Qy	601 IACLPNICTQPIPLLKOKNHISITMSQEQIASLLANAFCTFPRNNAKMSSESYSDI 660
Dd	601 IACLPNICTQPIPLLKOKNHISITMSQEQIASLLANAFCTFPRNNAKMSSESYSDI 660
Qy	661 NFNRLPEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 720
Dd	661 NFNRLPEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 720
Qy	721 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLPFTEVLHNE 780
Dd	721 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLPFTEVLHNE 780
Qy	781 CLIIITGTEQSYTGAETRYRWARSHEDSRDDQRRTTEIVAIDALHFRYLDQFVPE 840
Dd	781 CLIIITGTEQSYTGAETRYRWARSHEDSRDDQRRTTEIVAIDALHFRYLDQFVPE 840
Qy	841 KIRRELNKAYCGFLRPGVSSENLAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900
Dd	841 KIRRELNKAYCGFLRPGVSSENLAVATGNWCGGAFGGDARLKALIQILAAVAERDVVY 900
Qy	901 FTFGDSELMRDIYSMHTFLTERKLTGVGVYKLLRLYYNEECRNCSTPGPDIKLYPIYHA 960
Dd	901 FTFGDSELMRDIYSMHTFLTERKLTGVGVYKLLRLYYNEECRNCSTPGPDIKLYPIYHA 960
Qy	961 VESCTQTTNQPGQRTGA 977
Dd	961 VESCTQTTNQPGQRTGA 977

Db 61 GRAGQHRGSAATSLVFQKQITTSMDTKGKTVSESLHSEKNNNTREESMMSSVQKDNFY 120  
 QY 121 QHNMEKLENVSLGFDKSPVEKGTQYLKHQHTAAMCKWQNEGPHSRLLLESEPPAVTLVP 180  
 Db 121 QHNMEKLENVSLGFDKSPVEKGTQYLKHQHTAAMCKWQNEGPHSRLLLESEPPAVTLVP 180  
 QY 181 EQFSNANVDQSSPKDHDSTNSESNDNQFLTHVKLANAKQTMEDQEGREARSHQCKG 240  
 Db 181 EQFSNANVDQSSPKDHDSTNSESNDNQFLTHVKLANAKQTMEDQEGREARSHQCKG 240  
 QY 241 ACHPAACAGCQOEEDTVVSESLSDTSGEDVGTGLKNANRLNRQSSSLGNSPPFFKESE 300  
 Db 241 ACHPAACAGCQOEEDTVVSESLSDTSGEDVGTGLKNANRLNRQSSSLGNSPPFFKESE 300  
 QY 301 PESPMDVNSKNSCQDSEADEETSPGDEQDSSAQTANKPSRFPREADTELKRS 360  
 Db 301 PESPMDVNSKNSCQDSEADEETSPGDEQDSSAQTANKPSRFPREADTELKRS 360  
 QY 361 KGGEIRLHFQEGEGSRAGNDVNAKPGSTSSLANVECRNSKQHGKDSKITDHFMRVPK 420  
 Db 361 KGGEIRLHFQEGEGSRAGNDVNAKPGSTSSLANVECRNSKQHGKDSKITDHFMRVPK 420  
 QY 421 AEDKXKEQCEMKHQRTERKIPKIPPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 480  
 Db 421 AEDKXKEQCEMKHQRTERKIPKIPPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 480  
 QY 481 HTVTIRVDLLRIGEPKPPFTFKDLWKNKHVQMPCEQNLYPVEDENGERRAGSWELI 540  
 Db 481 HTVTIRVDLLRIGEPKPPFTFKDLWKNKHVQMPCEQNLYPVEDENGERRAGSWELI 540  
 QY 541 OTALLNRLTRPQNLKDAILKYNVAYSKWDFTALIDFWDKVLAEAAQHLQYSLPDMVK 600  
 Db 541 OTALLNRLTRPQNLKDAILKYNVAYSKWDFTALIDFWDKVLAEAAQHLQYSLPDMVK 600  
 QY 601 IALCLPNICTQPIPLKQKWHSTTSQEQIASILLANAFCTPPRNNAKMSYSSYPDI 660  
 Db 601 IALCLPNICTQPIPLKQKWHSTTSQEQIASILLANAFCTPPRNNAKMSYSSYPDI 660  
 QY 661 NFNRLEFGRSSRKPXKLTLCFYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKLTLRH 720  
 Db 661 NFNRLEFGRSSRKPXKLTLCFYFRRVTEKKPTGLVTFTRQSLDFPEWERCCKLTLRH 720  
 QY 721 VTYEGTIEGNGQMLQVDFANRFGVGGVTSAGLVQBEIRPLINPELIVSLFTEVLHNE 780  
 Db 721 VTYEGTIEGNGQMLQVDFANRFGVGGVTSAGLVQBEIRPLINPELIVSLFTEVLHNE 780  
 QY 781 CLITTEQVSEYTGVAETVWARSHEDESRDDWQRTTEIIVAIIDALHPRYLDQFVPE 840  
 Db 781 CLITTEQVSEYTGVAETVWARSHEDESRDDWQRTTEIIVAIIDALHPRYLDQFVPE 840  
 QY 841 KIRRELKAYCGFLRPGVSSENISAVATGNWGCAGFGGDARLKALIQILAAVAERDVY 900  
 Db 841 KIRRELKAYCGFLRPGVSSENISAVATGNWGCAGFGGDARLKALIQILAAVAERDVY 900  
 QY 901 FTGDSLMEDYISMHTFLERKLTGVEVYKLLRYNEECRCNCTPGPDIKLYPFIYH 960  
 Db 901 FTGDSLMEDYISMHTFLERKLTGVEVYKLLRYNEECRCNCTPGPDIKLYPFIYH 960  
 QY 961 VESCTQTTPQPGORTGA 977  
 Db 961 VESCTQTTPQPGORTGA 977

RESULT 5  
 AAE25630  
 ID AAE25630 standard; protein; 976 AA.  
 XX  
 AC AAE25630;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX  
 KW Human: poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme; ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic; cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity; Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac; cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy; antisense therapy.  
 OS Homo sapiens.  
 XX US6395543-B1.  
 XX 28-MAY-2002.  
 XX 23-FEB-2000; 2000US-00511507.  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 DR N-PSDB; AAD42082.  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.  
 PS Claim 3; Col 55-60; 77pp; English.  
 XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyzes release of ADP-ribose from an ADP-ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridization assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG

Sequence 976 AA;  
 Query Match 89.4%; Score 4635.5; DB 5; Length 976;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MSAGPCCECTKPRWDAAATSPPAASDARSFPGRRVLDKDAVPQVRVPPSSGSCAL 60  
 Db 1 MNAGPCCECTKATRWGAATTS-PAASDARSFPGRRVLDKDAHVPQVRVPPSSPACVP 59  
 QY 61 GRAGQHRGSAATSLVFQKQITTSMDTKGKTVSESLHSEKNNNTREESMMSSVQKDNFY 120  
 Db 60 GQAGQHRGSAATSLVFQKQITTSMDTKGKTVSESLHSEKNNNTREESMMSSVQKDNFY 119  
 QY 121 QHNMEKLENVSLGFDKSPVEKGTQYLKHQHTAAMCKWQNEGPHSRLLLESEPPAVTLVP 180  
 Db 120 QHNMEKLENVSLGFDKSPVEKGTQYLKHQHTAAMCKWQNEGPHSRLLLESEPPAVTLVP 179  
 QY 181 EQFSNANVDQSSPKDHDSTNSESNDNQFLTHVKLANAKQTMEDQEGREARSHQCKG 240  
 Db 180 EQFSNANVDQSSPKDHDSTNSESNDNQFLTHVKLANAKQTMEDQEGREARSHQCKG 239  
 QY 241 ACHPAACAGCQOEEDTVVSESLSDTSGEDVGTGLKNANRLNRQSSSLGNSPPFFKESE 300  
 Db 240 SCHPGEDCASCQOEDTVVVPKSPSLSDVGSDEVDGTGSKNKNKLRQESCLGNSPPFFKESE 299



QY 301 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSAQNTANKPSRFQPPREADTELEKRS360  
 DB 300 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSAQNTANKPSRFQPPREADTELEKRS359  
 QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKRGSTSSLVNVECRNSKHQGRKDSKITDHPMRVPK 420  
 DB 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPNTSSLVNVECRNSKHQGRKDSKITDHLMLRPLK 419  
 QY 421 AEDRKEQCEMKHQTERTKIPKYPHLSPPDKKMLGTPIEEMRMRPRCIGIRLPPLRPSAN 480  
 DB 420 AEDRKEQCEMKHQTERTKIPKYPHLSPPDKKMLGTPIEEMRMRPRCIGIRLPPLRPSAN 479  
 QY 481 HTVTIRVDLLRIGRVPKPPPTFKDLMDNKHVMPCSEONLYPVDENGERAGSRWELI 540  
 DB 480 HTVTIRVDLLRAGEVPKPPPTFKDLMDNKHVMPCSEONLYPVDENGERTAGSRWELI 539  
 QY 541 QTALLNRLTRPNQDLAILKYNVASKKWDFTALIDFWDKVLLEAQAHLQSIILPDMVK 600  
 DB 540 QTALLNKFTRPNQDLAILKYNVASKKWDFTALIDFWDKVLLEAQAHLQSIILPDMVK 599  
 QY 601 IALCLPNICTQPIPLLKQKWHISITMSQBIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
 DB 600 IALCLPNICTQPIPLLKQKWHISITMSQBIASLLANAFCTFPRRNAMKSEYSSYPDI 659  
 QY 661 NFNRLEFGSRSSRPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRI 720  
 DB 660 NFNRLEFGSRSSRPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRI 719  
 QY 721 VTYEGTIEGNGGMLQVDFANRFGVGVTSAGLVOEIRFLINPILNLSRLEFTEVLDRNE 780  
 DB 720 VTYEGTIEGNGGMLQVDFANRFGVGVTSAGLVOEIRFLINPILNLSRLEFTEVLDRNE 779  
 QY 781 CLIIITGTEQSYTGYAETRYRWARSHEDRSERDDWQRTTEIVAIDLHFRYRLDQFVPE 840  
 DB 780 CLIIITGTEQSYTGYAETRYRWARSHEDRSERDDWQRTTEIVAIDLHFRYRLDQFVPE 839  
 QY 841 KIRRELNAKAYCFLRPGVSSENLAVATGNWCGGAFGGDARUKALIQILAAVAERDVVY 900  
 DB 840 KMRRELNAKAYCFLRPGVSSENLAVATGNWCGGAFGGDARUKALIQILAAVAERDVVY 899  
 QY 901 FTFGSELMDIYSMHITFLTERKLTVEGVYKLLRYNEECRNCSTPGPDIKLYPIFYHA 960  
 DB 900 FTFGSELMDIYSMHITFLTERKLTVEGVYKLLRYNEECRNCSTPGPDIKLYPIFYHA 959  
 QY 961 VESCTQTNNQPGQRTG 976  
 DB 960 VESCAETADHSGQRTG 975  
 RESULT 6  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX AC AAU76021;  
 XX DT  
 XX DE  
 XX KW Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX KW Human; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX OS Homo sapiens.  
 XX PN US6337202-B1.  
 XX PD 08-JAN-2002.  
 XX PF 23-FEB-2000; 2000US-00511477.

XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI N-PSDB; ASK14932.  
 XX WPI; 2002-163240/21.  
 DR Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 DR which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 PT Claim 2; Col 55-60; 81pp; English.  
 PS The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 XX (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the human PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX Sequence 976 AA;  
 QY Query Match 89.4%; Score 4635.5; DB 5; Length 976;  
 DB Best Local Similarity 89.5%; Pred. No. 0;  
 QY Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 DB 1 MSAGPGCEPCTKRPWDAAATSPPAASDARSFPGRQRRVLDKADAPVFRVPPSSGAL 60  
 DB 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPGRQRRVLDKADAPVFRVPPSPACVP 59  
 QY 61 GRAGQHRGSATSLVPKQKTIITSMWDTKGIKTVESLSLKNNNTRESSMSVQKDNEY 120  
 DB 60 GQAGQHRGSATSLVPKQKTIITSMWDTKGIKTAESLSLKNNNTRESSMSVQKDNEY 119  
 QY 121 QHNMEKLENVSQLGFDKSPVEKGTOLKQHOTAAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
 DB 120 QHNVEKLVNVSQLSLDKSLTEKSTQYLNQHQTAAMCKWQNEGKHTEQELLESEPPQVTLVP 179  
 QY 181 EQFSNANVDQSPKDDHSDTNSSESRDNOQFLTHVKLANAKQTMEDQEGREARSHQKCK 240  
 DB 180 EQFSNANIDRSPQNDHSDTNSSESRDNOQFLTHVKLANAKQTMEDQEGREARSHQKCK 239  
 QY 241 ACHPAEACAGCQOEEDTVVSESPSLDTSQEDVGTGLKNAKLNROESSIGNSPPPEKSE 300  
 DB 240 SCHPCEDCASCQOEEDTVVSESPSLDTSQEDVGTGLKNAKLNROESSIGNSPPPEKSE 299  
 QY 301 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSAQNTANKPSRFQPPREADTELEKRS360  
 DB 300 PESPMVDVNSKNSCODSEADETSPGFDEQEDSSAQNTANKPSRFQPPREADTELEKRS359  
 QY 361 KGGEIRLHFQFEGGESRAGMNDVNAKRGSTSSLVNVECRNSKHQGRKDSKITDHPMRVPK 420  
 DB 360 KGGEVRLHFQFEGGESRTGMNDLNAKLPNTSSLVNVECRNSKHQGRKDSKITDHLMLRPLK 419  
 QY 421 AEDRKEQCEMKHQTERTKIPKYPHLSPPDKKMLGTPIEEMRMRPRCIGIRLPPLRPSAN 480  
 DB 420 AEDRKEQCEMKHQTERTKIPKYPHLSPPDKKMLGTPIEEMRMRPRCIGIRLPPLRPSAN 479  
 QY 481 HTVTIRVDLLRIGRVPKPPPTFKDLMDNKHVMPCSEONLYPVDENGERAGSRWELI 540

Db 480 HTVTRVLLRAGEVPKPPFTHYKDLNDKNDKVVMPCEQNLVPEDENGERTAGSRWELI 539  
 QY 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFDALDFWQKVLLEAEQAHLVQSIILPDMVK 600  
 Db 540 QTALLNKRTPQNLKDALIKYNVAYSKKWDFDALDFWQKVLLEAEQAHLVQSIILPDMVK 599  
 QY 601 TALCLPNICTQPIPLIKQNMHSITMSQBIASILLANAFCTFPRNNAKMSSEYSSYDI 660  
 Db 600 TALCLPNICTQPIPLIKQNMHSITMSQBIASILLANAFCTFPRNNAKMSSEYSSYDI 659  
 QY 661 NFNRLFEGRSRKPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 720  
 Db 660 NFNRLFEGRSRKPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 719  
 QY 721 VTYEGTIEGQGMQLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNE 780  
 Db 720 VTYEGTIEGQGMQLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNE 779  
 QY 781 CLITGTQYSEYTCYAEYRWARSHEDSERDDWQRTEITVAIDALHFRYLDQVPE 840  
 Db 780 CLITGTQYSEYTCYAEYRWARSHEDSERDDWQRTEITVAIDALHFRYLDQVPE 839  
 QY 841 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 900  
 Db 840 KIRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAVAERDVVY 899  
 QY 901 FTFGSELMDIYSMTFLTERKLTVEGYKLLRYNEECNRCSTPGPDIKLYPFIYHA 960  
 Db 900 FTFGSELMDIYSMTFLTERKLTVEGYKLLRYNEECNRCSTPGPDIKLYPFIYHA 959  
 QY 961 VESCTQTINQPGQRTG 976  
 Db 960 VESCAETADHSGQRTG 975

## RESULT 7

AAU76012  
 ID AAU76012 standard; protein; 976 AA.

XX AC AAU76012;  
 XX 08-MAY-2002 (first entry)  
 XX DE Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX Human; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX OS Homo sapiens.  
 XX PN US6333148-B1.  
 XX PD 25-DEC-2001.  
 XX PF 30-APR-1999; 99US-00302812.  
 XX PR 01-MAY-1999; 98US-0083768P.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX PI WPI; 2002-153820/20.  
 XX DR N-PSDB; ABK14494.  
 XX PT Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX PS Claim 3; Col 55-60; 80pp; English.

XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly (ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the human PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76015) of the  
 CC invention  
 XX Sequence 976 AA;

Query Match 89.4%; Score 4635.5; DB 5; Length 976;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MSAGPGCEPCTKRPRWDAAATSPPAASDARGSPGQRRLVDSKAPVQFRVPPSSGCAI 60  
 Db 1 MNAGPGCEPCTKATWGAATTS-PAASDARGSPGQRRLVDPKDAHVQFRVPPSPACVP 59  
 QY 61 GRAGQHRGASATSLVPKQKTIITSMWDTGKGTVESESLHSKNNNTRESMMSVQKDNFY 120  
 Db 60 GQAGQHRGASATSLVPKQKTIITSMWDTGKGTAESESLDSKNNNTRIEMSMSSVQKDNFY 119  
 QY 121 QHNMEKLENVQLGPKSPVEKGTQYLKQHOTAAKQNEGPHSERLLESPPAVTLVP 180  
 Db 120 QHNVEKLVNVSQSLDSLSEKSTQYLKQHOTAAKQNEGKHTQLLESPPQVTLVP 179  
 QY 181 EOPSNNANVDQSPKDDHSDTMSSESRDNQOFLTHVKLANAKOTMEDBQREARSQKCK 240  
 Db 180 EOPSNNANIDRSPQNDHSDTSEENRDNQOFLTHVKLANAKOTTEDEHAREAKSHQCKSK 239  
 QY 241 ACHPAEACAGCQOEETDVVSESPISDTCSEVDGTGLKNANRLNROESSLGNSSPPEKSE 300  
 Db 240 SCHPOEDCASCQODEIDVVPKSPISDVGSESDVGTGSKNDNKLIRQESCLGNSPPEKSE 299  
 QY 301 PESPMVDVNSKNSCODSEADETSPGFDEQESSAQATANKPSRFPQREADTELKRS 360  
 Db 300 PESPMVDVNSKNSCODSEADETSPGFDEQESSAQATANKPSRFOADADIEFKRYST 359  
 QY 361 XGGEIRLHFQEGGESRAGMNDVNAKPGSTSSLNVECRNSKQHGKDKSKITDHFMRVVK 420  
 Db 360 XGGEVRLHFQEGGESRAGMNDVNAKPGSTSSLNVECRNSKQHGKDKSKITDHLMLPK 419  
 QY 421 AEDKKEQCEMKHQRTERKIPKYPHLSPPKMLGTPIEEMRMPCGIRLPPRPSAN 480  
 Db 420 AEDRRKEQWETKHQRTERKIPKYPVPHLSPPDKMLGTPIEEMRMPCGIRLPPRPSAN 479  
 QY 481 HTVTRVLLRAGEVPKPPFTHYKDLNDKNDKVVMPCEQNLVPEDENGERAAGSRWELI 540  
 Db 480 HTVTRVLLRAGEVPKPPFTHYKDLNDKNDKVVMPCEQNLVPEDENGERTAGSRWELI 539  
 QY 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFDALDFWQKVLLEAEQAHLVQSIILPDMVK 600  
 Db 540 QTALLNKRTPQNLKDALIKYNVAYSKKWDFDALDFWQKVLLEAEQAHLVQSIILPDMVK 599  
 QY 601 TALCLPNICTQPIPLIKQNMHSITMSQBIASILLANAFCTFPRNNAKMSSEYSSYDI 660  
 Db 600 TALCLPNICTQPIPLIKQNMHSITMSQBIASILLANAFCTFPRNNAKMSSEYSSYDI 659  
 QY 661 NFNRLFEGRSRKPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 720  
 Db 660 NFNRLFEGRSRKPEKLTLCYFRVRVTEKPTGLVTFTRQSLDFPPEWERCEKLLTRLH 719  
 QY 721 VTYEGTIEGQGMQLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNE 780  
 Db 720 VTYEGTIEGQGMQLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRFLTEVLDHNE 779

QY 781 CLITGTEQYSEYTGATYTRWASHEDSRDDMQRTTEIIVADALHFRYLDQVPE 840  
 DB 780 CLITGTEQYSEYTGATYTRWASHEDSRDDCERCTEIVADALHFRYLDQVPE 839  
 QY 841 KIRRELKAYCGFLRPGVSSENISAVATGNGCGAFGSDARLKALIQILAAAARVDVY 900  
 DB 840 KIRRELKAYCGFLRPGVSSENISAVATGNGCGAFGSDARLKALIQILAAAARVDVY 899  
 QY 901 FTGDSSELMRDYISMHTFLTERKLTGVEYKLLRYNNEECRCNSTPGDILKYPFIYHA 960  
 DB 900 FTGDSSELMRDYISMHTFLTERKLTGVDYKLLRYNNEECRCNSTPGDILKYPFIYHA 959  
 QY 961 VESCTQTNQPGQRTG 976  
 DB 960 VESCAETADHSGQRTG 975

RESULT 8  
 ABG72279  
 ID ABG72279 standard; protein; 976 AA.  
 XX  
 AC ABG72279;  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX  
 KW Human; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; neotropic; antiparkinsonian; cardiac; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002132328-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 09-OCT-2001; 2001US-00973451.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEU/) AME J.  
 PA (LINW/) LIN W.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 WP: 2003-155895/15.  
 DR N-PSDB; ABX14478.  
 XX

New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose) glycohydrolase activity, for preventing, treating, or ameliorating a disease condition, e.g. neoplastic disorder, myocardial infarction or vascular stroke.

Claim 28; Fig 16; 86pp; English.

The present invention relates to the isolation of poly(ADP-ribose) glycohydrolase (PARG) from several species, and the polynucleotide sequences encoding them. Methods for inhibiting PARG expression or overexpressing PARG are also disclosed. PARG is involved in the cellular response to DNA damage, and is associated with the body's response to neoplastic disorder inducing agents and oxidative stress. The polynucleotide sequences encoding PARG and PARG modulators are useful for preventing, treating, or ameliorating diseases such as neoplastic disorders, myocardial infarction, vascular stroke, neurodegenerative

CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's disease), inborn genetic errors, reperfusion following ischaemia, aging, CC and neurotoxicity. The polynucleotide sequences are also useful in gene CC therapy. The methods are useful for identifying an agent that modulates CC PARG activity, identifying a mutant PARG allele in an individual, or CC screening candidate molecules for PARG modulating activity. The present CC sequence represents human PARG enzyme

XX  
 SQ Sequence 976 AA;  
 Query Match 89.4%; Score 4635.5; DB 6; Length 976;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY 1 MSAGPGCEPCTKRPRWDAATSPPAASDARSFPGRRVLDSDKAPVQFRVPPSSSGCAL 60  
 DB 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFPGRRVLDSDKAPVQFRVPPSSPACVP 59  
 QY 61 GRAGQHRGSAATSLVFYKQKTTISWMDTKGKTVESESLHSENNNTREESNMSSVQKDNFY 120  
 DB 60 GQAGQHRGSAATSLVFYKQKTTISWMDTKGKTVESESLHSENNNTREESNMSSVQKDNFY 119  
 QY 121 QHNMEKLENVSQLGDPKSPVEKGTQYLKQHOTAAKCKWQNEGPHSRLLSESEPPAVTLVP 180  
 DB 120 QHNVEKLENVSQLGDPKSPVEKGTQYLKQHOTAAKCKWQNEGPHSRLLSESEPPAVTLVP 179  
 QY 181 EQFSNANVDQSSPKDDHSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHQCKGK 240  
 DB 180 EQFSNANIDRSQNDQSDHSDTNSSESDNQFLTHVKLANAKQTMEDQEGREARSHQCKGK 239  
 QY 241 ACPAECACAGCQEEFTDVVSESLSDTSGEDVGTGLKNAFLNRORSSLCNSPPEKESE 300  
 DB 240 SCHPGSDCASCQDEIDVVPKSLDVGSDEVTGSKNDKLLRQESCLNSPPEKESE 299  
 QY 301 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSAQATANKPRFPQPREADTELKRESSA 360  
 DB 300 PESPMVDVNSKNSCQDSEADETSPGFDEQEDSSAQATANKPRFPQPREADTELKRESSA 359  
 QY 361 KGGEIRLHFQFEGESRAGMNVNAKPGSTSSLVNVECRNSKQHGKDKSKITDHFMRVVK 420  
 DB 360 KGGEVRLHFQFEGESRGTGMDNLAKLPGNISLVNVECRNSKQHGKDKSKITDHLRLPK 419  
 QY 421 AEDKRXEQCEMKHQRTERKIPKIPHLSPDKKWLGTPIEMERMPRCGRIRLPLRPSAN 480  
 DB 420 AEDKRXEQCEMKHQRTERKIPKIPHLSPDKKWLGTPIEMERMPRCGRIRLPLRPSAN 479  
 QY 481 HTVTIRVDLLRIGEVKPKPTPTFKDLWKNKHVWPCSEQNLVPEVDENGERRAGSRWELI 540  
 DB 480 HTVTIRVDLLRIGEVKPKPTPTFKDLWKNKHVWPCSEQNLVPEVDENGERRAGSRWELI 539  
 QY 541 QTALLNRLTRPQNLKDAILLKYNVAYSKKWDFTALIDFWDKVLEBAQAHLIYQSILPDMVK 600  
 DB 540 QTALLNRLTRPQNLKDAILLKYNVAYSKKWDFTALIDFWDKVLEBAQAHLIYQSILPDMVK 599  
 QY 601 IALCLPNICTQPIPLLKQKNHSHITWSQOIASLLANAFCTEPPRNAMKMSYSSVPDI 660  
 DB 600 IALCLPNICTQPIPLLKQKNHSHITWSQOIASLLANAFCTEPPRNAMKMSYSSVPDI 659  
 QY 661 NFNRLEFGRSSRKPEKLTLCYFRVTEKPKTGLVTFTRQSLDEDFPEWERCETLRLH 720  
 DB 660 NFNRLEFGRSSRKPEKLTLCYFRVTEKPKTGLVTFTRQSLDEDFPEWERCETLRLH 719  
 QY 721 VTYEGTIEGNGQMLQVDFANRFGVGGVTSAGLVQBEIRFLINPELISRLFTFVLDHNE 780  
 DB 720 VTYEGTIEGNGQMLQVDFANRFGVGGVTSAGLVQBEIRFLINPELISRLFTFVLDHNE 779  
 QY 781 CLITGTEQYSEYTGATYTRWASHEDSRDDMQRTTEIIVADALHFRYLDQVPE 840  
 DB 780 CLITGTEQYSEYTGATYTRWASHEDSRDDCERCTEIVADALHFRYLDQVPE 839  
 QY 841 KIRRELKAYCGFLRPGVSSENISAVATGNGCGAFGSDARLKALIQILAAAARVDVY 900  
 DB 840 KIRRELKAYCGFLRPGVSSENISAVATGNGCGAFGSDARLKALIQILAAAARVDVY 899

QY 901 FTFGDSELMRDYIMTIFLTERKLTGVGVYKLLRYNEECRCNSTPGDDIKLYPIYHA 960  
 Db 900 FTFGDSELMRDYIMHIFLTERKLTGVGVYKLLRYNEECRCNSTPGDDIKLYPIYHA 959  
 QY 961 VESCTQTTNPGQRTG 976  
 Db 960 VESCAETADHSGQRTG 975  
 RESULT 9  
 AAE25631  
 ID AAE25631 standard; protein; 968 AA.  
 XX  
 AC AAE25631;  
 XX 04-NOV-2002 (first entry)  
 XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 DE  
 DE Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; nootropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Mus musculus.  
 XX  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42083.  
 DR  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 PS Claim 3; Col 63-68; 77pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG  
 XX  
 SQ Sequence 968 AA;

Query Match 83.6%; Score 4332.5; DB 5; Length 968;  
 Best Local Similarity 84.0%; Pred. No. 0;  
 Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;  
 QY 1 MSAGPGCEPCTKRPDAAATSPPAASDARSPPGQRRLVDSKADAPVQVRPVPSSSGCAL 60

Db 1 MSAGPGWBECKT-AWGAAAGTSAPTASDSRSPPGQRRLVDPKADVPQVRPVPSSPACVS 59  
 QY 61 GRAGQHRGSATSFLVKQKQKTIITSWMDTKGIKTVESESLHSEKNNNTREESNMSSVQKDNFY 120  
 Db 60 GRAGPHRGNATSFVKQKTIITWMDTKGPKTAESE--SKENNTRIDSMSSVQKDNFY 116  
 QY 121 QHNMEKLENVSQIGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSERLLESEPPATVLVP 180  
 Db 117 PHKVEKLENVPQNLNLDKSTFKSQYLNQOQTASVCKWQNEGKHAEQLLASEPPAGTPLP 176  
 QY 181 EQFSNANVDQSSPKDDHSDTNSSESRDNOQFTTHVKLANAKOTMEDEQGREARSHOKCGK 240  
 Db 177 KQLSNANIGQSHDTHSDTDEHEDRDNOQFTPIKLANTKPTVGDGQ--ARSNCKCSG 233  
 QY 241 ACHPAEACAGCQOEERTDVVSEPLSDTSGSEDTGTLKNAANRLNRQESSLGNPPPEKESE 300  
 Db 234 SRQSVKDCCTGCOQEEVDVLPESPLSDVGAEDIGTGPKNDNKLTDQESSLGDSPPEKESE 293  
 QY 301 PESPMDVNSKNSCODSEADEETSFGFDEQDESSAQATANKFSRFPQPREADTELKRSAS 360  
 Db 294 PESPMDVNSNRNSCODSEADEETSPVDEQDDRSS--QTANKLSQCAQREADGDLKRYLT 352  
 QY 361 KGEIRLHFQREGGSRAGMNDVNAKRGSTSSLANVECHSKOGRKDSKIDTHFMKRVK 420  
 Db 353 KGEVRLHFQFE--GENNAGTSDLNAPSGNSSSLNVECHSKQHGKRDSDKIDTHFMKRSK 411  
 QY 421 AEDRKEQCEMKHQHRTERKIPKYIPPHLSPDKKLTGTPIEEMRMPCRGIRLPLPSAN 480  
 Db 412 SEDRKEQCEVHRHQHRTERKIPKYIPPNLPPPEKKWLTGTPIEEMRMKPCRCGHLPLSPASAS 471  
 QY 481 HTVTIRVLLRLRIGEVKPPPHFKDLWNKHVKMPCSEONLYPVEDENGERRAGSEWELI 540  
 Db 472 HTVTIRVLLRAGEVPKPFPTHYKDLWNKHVKMPCSEONLYPVEDENGERRAGSEWELI 531  
 QY 541 OTALLNRLTRPONLADAILKYNVAYSKKWDFTALIDFWDKVLSEAEAOHLYOSILPDMVK 600  
 Db 532 QTALLNKTRPONLADAILKYNVAYSKKWDFTALVDFWDKVLSEAEAOHLYOSILPDMVK 591  
 QY 601 TALCLPNICTQPIPLLKQNMNHSITMSQEQIASLLANAFCTTFFPRRNAMKSEYSYDPI 660  
 Db 592 TALCLPNICTQPIPLLKQNMNHSITMSQEQIASLLANAFCTTFFPRRNAMKSEYSYDPI 651  
 QY 661 NNRLFEGRSRKPEKLTLCYFRRVTEKPTGLVTTQSLDEDPPEWERCEKLLTRLH 720  
 Db 652 NNRLFEGRSRKPEKLTLCYFRRVTEKPTGLVTTQSLDEDPPEWERCEKLLTRLH 711  
 QY 721 VTYEGTIEGNGQGLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLDHNE 780  
 Db 712 VTYEGTIEGNGRGLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLDHNE 771  
 QY 781 CLITITGTOYSEYTGAYETIYWARSHEDRSDDWQRRTTEIVAIDALHFRYLDQFVPE 840  
 Db 772 CLITITGTOYSEYTGAYETIYWARSHEDSGSEKDWQRRETEIVAIDALHFRYLDQFVPE 831  
 QY 841 KIRRELKAYCGELRPGVSSNLSAVATGNWCGGAGGDARLKALIQILAAVAERDVVY 900  
 Db 832 KVRRELKAYCGELRPGVSSNLSAVATGNWCGGAGGDARLKALIQILAAVAERDVVY 891  
 QY 901 FTFGDSELMRDYIMHIFLTERKLTGVGVYKLLRYNEECRCNSTPGDDIKLYPIYHA 960  
 Db 892 FTFGDSELMRDYIMHIFLTERKLDVGKVKLLRYNEECRCNSTPGDDIKLYPIYHA 951  
 QY 961 VESCTQTTNPGQRTG 976  
 Db 952 VESSAETDMPGQKAG 967

RESULT 10  
 AAU76022 standard; protein; 968 AA.  
 XX  
 AC AAU76022;

XX DT 08-MAY-2002 (first entry)

XX DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;

XX KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;

XX KW leukaemia; lymphoma; melanoma; sarcoma; teratocarcinoma;

XX KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;

XX KW neurodegenerative disease; neurological disorder; Alzheimer's disease;

XX KW Huntington's disease; Parkinson's disease.

XX OS Mus musculus.

XX PN USG337202-B1.

XX PD 08-JAN-2002.

XX PF 23-FEB-2000; 2000US-00511477.

XX PR 01-MAY-1998; 98US-0083768P.

XX PR 30-APR-1999; 99US-00302812.

XX PA (KENT) UNIV KENTUCKY RES FOUND.

XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;

XX DR WPI; 2002-163240/21.

XX DR N-PSDB; ABK14933.

XX PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein

XX PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for

XX PT treating neoplastic and neurological disorders, heart attack and stroke.

XX PS Claim 2; Col 63-70; 81pp; English.

XX CC The present invention relates to a new poly(ADP-ribose) glycohydrolase

XX CC (PARG) protein which catalyses release of ADP-ribose from an ADP

XX CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the

XX CC invention is useful for generating antibodies and can be inhibited or

XX CC activated for diagnosing and treating neoplastic disorders such as

XX CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,

XX CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following

XX CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological

XX CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,

XX CC and related conditions. PARG levels may be enhanced to suppress DNA

XX CC repair and increase the cell's susceptibility to chemotherapy drugs.

XX CC Antagonists of PARG are administered to treat or prevent neoplastic

XX CC disorders. The present amino acid sequence represents the mouse PARG

XX CC protein of the invention. This protein is one of several PARG proteins

XX CC (AAU76020-AAU76024) of the invention

XX SQ Sequence 968 AA;

Query Match 83.6%; Score 4332.5; DB 5; Length 968;

Best Local Similarity 84.0%; Pred. No. 0;

Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGCGEPTCKRPWDAAATSPPAASDARSPGQRVLDKAPVQFRVPPSSGCGAL 60

DB 1 MSAGGWEPTCK-ARWGAAGTSAPTASDSRSPGQRVLDKAPVQFRVPPSSPACVS 59

QY 61 GRAGHRSATSLVFKQTTTSMWTKGKTVESSLHSKNNNTRESMSSVQKDNFY 120

DB 60 GRAGPHRGNATSFVFKQTTTSMWTKGPKTAESE---SKNNNTRESMSSVQKDNFY 116

QY 121 QHNMEKLENVSLQGDGKSPVEKGTQYKQHOTAAACKWQNEGPHSERLLESEPPAVTLVP 180

DB 117 PHKVKLENVLPQNLIDKSPTEKSSQYLNQQTASVCKWQNEGKHAEQLLASEPPAGTLP 176

QY 181 EQFSNANVDQSPKDDHSDTNSSESRDQQLFTHVKLANAKQTMEDQEGREARSHQKCGK 240

DB 177 KQLSNANIGQSPHTDDHSDTDEEDRDNQQLFTPIKLANTKFTVGDGQ---ARSNCKCSG 233

QY 241 ACHPAEACAGCQOEETDVVSESPLSDTSGSEVDVGTGLKNAKLNQESSLGNSPPEKSE 300

DB 234 SRQSVKDCQCEEVDLVPESPLSDVGAEDIGTPKNDKNTLQESSLSDGSPPEKSE 293

QY 301 PESMDVDNKNKSCQDSEADBTSPGFDEQEDSSAQTANKPSPFQPREADTELKRRSA 360

DB 294 PESMDVDNKNKSCQDSEADBTSPGFDEQEDSSAQTANKPSPFQPREADTELKRRSA 352

QY 361 KGSIRLHFQEGESRAGMNDVNAKPGSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420

DB 353 KGSVRLHFQFE-GENNAGTSDLNKPSGNSSSLNVECRNSKQGRKDSKITDHFMRISK 411

QY 421 AEDKREKQCEMKHQRTERKPKYIPPHLSPDKWLGTPIDEMRMPCRGILPPLRPSAN 480

DB 412 SEDRRKEQCEVRHQRTERKPKYIPPHLSPDKWLGTPIDEMRMPCRGILPPLRPSAN 471

QY 481 HTVTIRVDLLRIGEVPPFPFTHFKDLMDNKHVMPCSEQNLYPVEDENGERAGSRWELI 540

DB 472 HTVTIRVDLLRIGEVPPFPFTHFKDLMDNKHVMPCSEQNLYPVEDENGERAGSRWELI 531

QY 541 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFITALIDFWDKVLSEAEQAHLQSLPDMVK 600

DB 532 QTALLNRLTRPQNLKDALIKYNVAYSKKWDFITALIDFWDKVLSEAEQAHLQSLPDMVK 591

QY 601 IALCLPNICTQPIPLLKQKNHSTMSQBOIASLLANAFCTFPRRNKMKSESYSPDI 660

DB 592 IALCLPNICTQPIPLLKQKNHSTMSQBOIASLLANAFCTFPRRNKMKSESYSPDI 651

QY 661 NFNRLFEGRSRPEKLTLCFYPRRVTEKPKTGLVTFTRQSLDFEPWERCCKLLTRLH 720

DB 652 NFNRLFEGRSRPEKLTLCFYPRRVTEKPKTGLVTFTRQSLDFEPWERCCKPLTRLH 711

QY 721 VTYEGTTEGNGQMLQVDFANRFVGGGVTSAGLVOEIRFELINDELIVSRILFTEVLHDNE 780

DB 712 VTYEGTTEGNGQMLQVDFANRFVGGGVTSAGLVOEIRFELINDELIVSRILFTEVLHDNE 771

QY 781 CLIIITGTEQSEYTGATVETRWARSHEDRSDQRTTETIVDAIDALHFRYLDQFVPE 840

DB 772 CLIIITGTEQSEYTGATVETRWARSHEDRSDQRTTETIVDAIDALHFRYLDQFVPE 831

QY 841 KIRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDAPLAKALIQLAAVAERDVVY 900

DB 832 KVRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDAPLAKALIQLAAVAERDVVY 891

QY 901 FTFGDSLEMDIYSNMHTFLTERKLTVCSEVYKLLRYNEECNCGSTPGPDIKLYPFIYHA 960

DB 892 FTFGDSLEMDIYSNMHTFLTERKLTVCSEVYKLLRYNEECNCGSTPGPDIKLYPFIYHA 951

QY 961 VESCTQTNNOPGORTG 976

DB 952 VESSAETTDMEGQKAG 967

RESULT 11

AAU76013

ID AAU76013 standard; protein; 968 AA.

XX AAU76013;

XX 08-MAY-2002 (first entry)

XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.

XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;

XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;

XX inherited genetic disease; myocardial infarction; vascular stroke; aging;

XX neurodegeneration; Huntington's disease; Parkinson's disease;

XX Alzheimer's disease; neurotoxicity.

XX Mus musculus.

XX USG333148-B1.

XX PD 25-DEC-2001.  
 XX PF 30-APR-1999; 99US-00302812.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14495.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX Claim 3; Col 63-68; 80pp; English.  
 XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the mouse PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX SQ Sequence 968 AA;  
 Query Match 83.6%; Score 4332.5; DB 5; Length 968;  
 Best Local Similarity 84.0%; Pred. No. 0;  
 Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;  
 QY 1 MSGAGCECTKPRWDAAATSPASDARSFGRORRVLDSDKADVPQVRPSSGGCAL 60  
 DB 1 MSGAGWEPTK-ARWGAAGTSAFTASDRSPGRORRVLDKADVPQVRPSSPACVS 59  
 QY 61 GRAGQHRGATSILVFQKQITSMMDTKG:KTVSESLHSKENNTRRESMSSVQKDNFY 120  
 DB 60 GRAGPHRGNATSVFQKQITITWMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116  
 QY 121 QINMEKLENVSQLGFDKSPVEKGTQYLKQHTAAMCKWQNEGPHSRLLLESPPAVTLVP 180  
 DB 117 PHKVEKLENVQNLNDKSPTEKSSQYLNQOQTASVCKWQNEGKHARQLLASBPAGTLP 176  
 QY 181 EQFSNANVDQSPKDDHSDTNEESRDNQFTLVKLANAKOTMEDEQGRARSHOKCKG 240  
 DB 177 KQLSNANTQSPHTDDHSDTDEEDNDNQFTPIKLANTKTVGDQ---ARSNCKSG 233  
 QY 241 ACHPAEACAGCQOEEETDVVSESPLSDTGSDEVTGLKNANRLNRQESSLGNPPFEKSE 300  
 DB 234 SRQSVKDCGTCQOEEVDVLPESPLSDVGAEDITGPKNDKLTGQESSLGDSPFEKSE 293  
 QY 301 PESPMDVNSKNSQDSEADETSPGDEQESSAQTANKPSRQPREADTELKRSRA 360  
 DB 294 PESPMDVNSRNSQDSEADETSPVDFEQDRSS--QTANKLSSCOAREADGLRKYLT 352  
 QY 361 KGGEIRLHFOFEGEGRAGNDVNAKPGSTSSLNVECRNSKQHGKDSKITDHFMRVPK 420  
 DB 353 KGSEVRLHFOFE--GENNATSDLNNAKPSGNSSSLNVECRSSKQHGKDSKITDHFMRISK 411  
 QY 421 AEDKRKEQCEMKHQRTERKIPKYPHLLSPDKWLGTPTEEMRMPCRGIRLPLRPSAN 480  
 DB 412 SEDRRKEQCEVHRQTERKIPKYPENLPPEKKWLGTPTEEMKWPCCGHLPLRPSAS 471  
 QY 481 HTVTIRVDLLRIGEVKPPPTTHYKOLWDNKNHVKMPCSEQNLVPVEDENGERTAGSRWELI 540

DB 472 HTVTIRVDLLRAGEVPPKPPPTTHYKOLWDNKNHVKMPCSEQNLVPVEDENGERTAGSRWELI 531  
 QY 541 QTALLNRLTRPONLKDAILKYNVAYSKKWDFTALIDFWDKVLLEAAQHLQSIILPDMVK 600  
 DB 532 QTALLNKFTRPONLKDAILKYNVAYSKKWDFTALVDFWDKVLLEAAQHLQSIILPDMVK 591  
 QY 601 IALCLPNICTQPIPLLKQKQNSHITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 660  
 DB 592 IALCLPNICTQPIPLLKQKQNSHITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651  
 QY 661 NFNRLFEGRSSRKPCKLKTLCYFRVTKKPTGLVTFTRQSLDDPPEWERCCKLTLRLH 720  
 DB 652 NFNRLFEGRSSRKPCKLKTLCYFRVTKKPTGLVTFTRQSLDDPPEWERCCKLTLRLH 711  
 QY 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGIYVDEIRFLINPELIVSRFLTEVLIDHNE 780  
 DB 712 VTYEGTIEGNGRMLQVDFANRFVGGVTSAGIYVDEIRFLINPELIVSRFLTEVLIDHNE 771  
 QY 781 CLITGTEQYSEYTGVAETIRWARSHEDRSERDDQRRTEITVAIDALHFRYLDQFVPE 840  
 DB 772 CLITGTEQYSEYTGVAETIRWARSHEDSGEXDDQRRCTEITVAIDALHFRYLDQFVPE 831  
 QY 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLQALQIILAAAVERDVTY 900  
 DB 832 KYRRELKAYCGFLRPGVSENLAVATGNWCGAFGGDARLQALQIILAAAVERDVTY 891  
 QY 901 FFFGSELMDIYSMHTFLTERKLTGVEYKLLRYNEECRCNSTPGPDIKLYPIYHA 960  
 DB 892 FFFGSELMDIYSMHTFLTERKLTGVEYKLLRYNEECRCNSTPGPDIKLYPIYHA 951  
 QY 961 VESCTQTTNQPCQRTG 976  
 DB 952 VESSAETTDMPQCKAG 967  
 RESULT 12  
 ABG72280  
 ID ABG72280 standard; protein; 968 AA.  
 XX AC ABG72280;  
 XX DT 13-MAR-2003 (first entry)  
 XX DE Mmurine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX KW Murine: poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; neurotropic; antiparkinsonian; cardiac;  
 KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
 XX OS Mus musculus.  
 XX PN US2002132328-A1.  
 XX PD 19-SEP-2002.  
 XX PF 09-OCT-2001; 2001US-00973451.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (JACO/) JACOBSON M K.  
 XX PA (JACO/) JACOBSON B L.  
 XX PA (AMEJ/) AME J.  
 XX PA (LINW/) LIN W.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX

DR WPI: 2003-155895/15.  
DR N-FSDB; ABX14479.  
PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
PT glycohydrolase activity, for preventing, treating, or ameliorating a  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
PT vascular stroke.  
XX  
PS Claim 28; Fig 16; 86pp; English.  
XX  
CC The present invention relates to the isolation of poly(ADP-ribose)  
CC glycohydrolase (PARG) from several species, and the polynucleotide  
CC sequences encoding them. Methods for inhibiting PARG expression or  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
CC response to DNA damage, and is associated with the body's response to  
CC neoplastic disorder inducing agents and oxidative stress. The  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
CC preventing, treating, or ameliorating diseases such as neoplastic  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
CC therapy. The methods are useful for identifying an agent that modulates  
CC PARG activity, identifying a mutant PARG allele in an individual, or  
CC screening candidate molecules for PARG modulating activity. The present  
CC sequence represents murine PARG enzyme  
XX  
SQ Sequence 968 AA;

Query Match 83.6%; Score 4332.5; DB 6; Length 968;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGGCEPCTKPRWDAAATSPPAASDARSFPGRORVLDKAPVQFRVPPSSGAL 60  
DB 1 MSAGGCEPCTKPRWDAAATSPPAASDARSFPGRORVLDKAPVQFRVPPSSGAL 60  
QY 61 GRAGHGRSATSILVPKQKTIISWMDTKIGKTVESLSHKNNNTRESMSSVQKNFY 120  
DB 60 GRAGHGRNATSFVKQKTIITWMDTKGPKTAESE---SKNNNTRESMSSVQKNFY 116  
QY 121 QHNMEKLENVSQLGFDKSPVEKGTQYQLKHQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
DB 117 PHKVEKLENVQLNLDKSPTEKSKSYLNQQTASVCKWQNEGKHAQLLASEPPAGTFLP 176  
QY 181 EOPSNANVDQSPPKDDHSDTNSERSRDNQOFLTHVKLANAKQWDEQREARSHQKCG 240  
DB 177 KQLSNANTGQSPHTDDHSDTDEEDRDNQOFLTPIKLANTKPTVGDQ---ARSNCKSG 233  
QY 241 ACHPAEACAGCQOEBTDVVSPLSDTSGEDVGTGLKNANLNQESSLGNSPPFKESE 300  
DB 234 SRQSVKDTGCGQOEBVDVLPESPJSDVGAELIGTGPKNDKLTGQESSLGDSPPFKESE 293  
QY 301 PESPMVDNKNKSCQDSADSETSPGDEQSDSSAQNTANKPSRFPQREADTELRKGS 360  
DB 294 PESPMVDNKNKSCQDSADSETSPVFEQDRESS-QTANKLSSCQAREADGDLRKRYLT 352  
QY 361 KGEIRLHFQEGESRAGMDVNAKPGTSSLNVECRNSKHGRKDKSKITDHFMRVPK 420  
DB 353 KGEVRLHFQFE-GENNAGTSDLNKPSGNSLNVESCRSKQHGKRSKITDHFMRLSK 411  
QY 421 AEDKEKEQCEMKHQRTKPKYIPPHLSPOKMWLTGPIEMRMMPRCGILPLRPSAN 480  
DB 412 SEDREKEQCEVHQRTKPKYIPNLPPEKMWLTGPIEMRMMPRCGILPLRPSAS 471  
QY 481 HTVTRVLLRIGVVPKPPFTFKDLNDKHKVMPCCSQNLYPVEDENGERRAAGSRWELI 540  
DB 472 HTVTRVLLRIGVVPKPPFTFKDLNDKHKVMPCCSQNLYPVEDENGERRAAGSRWELI 531  
QY 541 OTALNRLTRPNLKDAILKYNVASKKWDFTALDQWVKLEAEAOHLQSLIPDMVK 600  
DB 532 OTALNRLTRPNLKDAILKYNVASKKWDFTALDQWVKLEAEAOHLQSLIPDMVK 591

QY 601 IALCLPNICTQPIPLLKOKMNHSTMSQEOIASILANAAFFCTFPRRNAMKSEYSSYPDI 660  
DB 592 IALCLPNICTQPIPLLKOKMNHSTMSQEOIASILANAAFFCTFPRRNAMKSEYSSYPDI 651  
QY 661 NFNRLEFGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLDFPEWECCKLLRLH 720  
DB 652 NFNRLEFGRSSRKPEKLTILFCYFRRVTEKKPTGLVTFTRQSLDFPEWECCKPLTRLH 711  
QY 721 VTYEGTTEGNGQMLQVDFANRFVGGGVTSAGLVQEEIRFLINELIVSRLETFVLDHNE 780  
DB 712 VTYEGTTEGNGRGLQVDFANRFVGGGVTCAGLVQEEIRFLINELIVSRLETFVLDHNE 771  
QY 781 CLITGTGEQYSEYGYAETRWARSHEDSRDQWRRTTEIIVADALHFRYRLDQFVPE 840  
DB 772 CLITGTGEQYSEYGYAETRWARSHEDSGEKDQWRCTEIVADALHFRYRLDQFVPE 831  
QY 841 KIRRELNAKCYGFLRPGVSSSENLSAVATGNWCGAPGADARLKALIQILAAAVERDVVY 900  
DB 832 KVRRELNAKCYGFLRPGVPSSENLSAVATGNWCGAFGGADARLKALIQILAAAVERDVVY 891  
QY 901 FTFGDSSELMRDIIYSMHTFLTERKLTVCGEVVKLLRLRYNEECNCGSTPGPDIKLYPFIYHA 960  
DB 892 FTFGDSSELMRDIIYSMHTFLTERKLDVGKVKLLRLRYNEECNCGSTPGPDIKLYPFIYHA 951  
QY 961 VESCTOTTNPGQRTG 976  
DB 952 VESSAETDMEGQKAG 967

RESULT 13  
ABB59491  
ID ABB59491 standard; protein; 768 AA.  
XX ABB59491;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 5265.  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI: 2001-656860/75.  
XX N-FSDB; ABL03594.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-



CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from Wipo at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 768 AA;

Query Match 20.5%; Score 1063.5; DB 4; Length 768;  
 Best Local Similarity 41.5%; Pred. No. 6.2e-77;  
 Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

QY 417 RVPKAEKKEQCEMKHQRTERKIPKYPHPL--SPDKKWLGPPEM--RRMPCGIRL 472  
 DB 45 RNSKSPDGGISIEETEE-----PENLANSLLDSWRGVSMERHNRNQPFELENL 94  
 QY 473 PPLRPSANTVTVIRVDLLRIGEVKPPFTHFKDLWNKHKVMPKSCQNLVYVDENGERA 532  
 DB 95 PPVTAGNLHRVMYQ---LPFIRETP--PRPKSPGKWDSEHVRFLCPAPEKYPRENPDGTT 150  
 QY 533 AGSRWELIQTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLEAEQAHLQY 592  
 DB 151 IDFRWEMIERALLQPIKTCBELQAAIISNTYRQWHLRALHQLLDELDESETRVFE 210  
 QY 593 SILPDMVKIALCLPNICTOPILPKQKMNHSITMSQEQIASLLANAFCTFPRRNA--KMK 651  
 DB 211 DLLPRIIRLALRLPDLIQSPVPLKHKHKNASLSLQSQIISCLLANAFLCTFPRNTLKK 270  
 QY 652 SEYSSYPDINFNRLPGRSSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--L 703  
 DB 271 SEYSTPDPINFNRLYQSTGPAVLEKLCIMHYFRRVCPPTERDASNNVPTGVVTFVRSGLP 330  
 QY 704 EDFPEWERCEKLL--TRLHVTVEGTIEGNGQGMLOVDFAHRFVGGVTSAGLVQVEIRFL 761  
 DB 331 EHLIDWSQAAPLGDVPLHVDVDAEGTIEGIGLLQVDFANKVLGGVGHGCVQVEIRFV 390  
 QY 762 INPELIVSRFLTEVLHDNECLITGTQEYSEYTGVAETVYRWARSHDRSERRDWRRTTE 821  
 DB 391 ICPELLVGLKFTCELRPFALVMLGAERYSNITGYAGSFEWSGNFEDSPTRDSSGRQTA 450  
 QY 822 IVALDALHFRYLDQVPEKIRRELKAYCGFLR-----PGVSSENLSAVATGNWCGGA 875  
 DB 451 IVALDALHFAQSHHQYREDLMERELNKAYIGFVHMVMTTPPG-----VATGNWCGGA 502  
 QY 876 FGDDARLKALIQILAAVAERDVVYTFGDSSELMDIYSMTHTFLTERKLTVCGEVYKLLLR 935

RESULT 14  
 AA25632  
 ID AA25632 standard; protein; 768 AA.  
 AC AA25632;  
 XX  
 XX 04-NOV-2002 (first entry)  
 DT  
 DE Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 KW Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytosolic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy; fruit fly.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 XX US6395543-B1.  
 XX  
 XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42084.  
 DR  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 PS Example 13; Col 71-76; 77pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (PARG)  
 CC which catalyzes release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present invention is fruit fly PARG  
 XX  
 SQ Sequence 768 AA;

Query Match 20.5%; Score 1063.5; DB 5; Length 768;  
 Best Local Similarity 41.5%; Pred. No. 6.2e-77;  
 Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

QY 417 RVPKAEKKEQCEMKHQRTERKIPKYPHPL--SPDKKWLGPPEM--RRMPCGIRL 472  
 DB 45 RNSKSPDGGISIEETEE-----PENLANSLLDSWRGVSMERHNRNQPFELENL 94  
 QY 473 PPLRPSANTVTVIRVDLLRIGEVKPPFTHFKDLWNKHKVMPKSCQNLVYVDENGERA 532  
 DB 95 PPVTAGNLHRVMYQ---LPFIRETP--PRPKSPGKWDSEHVRFLCPAPEKYPRENPDGTT 150  
 QY 533 AGSRWELIQTALLNRLTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLEAEQAHLQY 592  
 DB 151 IDFRWEMIERALLQPIKTCBELQAAIISNTYRQWHLRALHQLLDELDESETRVFE 210  
 QY 593 SILPDMVKIALCLPNICTOPILPKQKMNHSITMSQEQIASLLANAFCTFPRRNA--KMK 651  
 DB 211 DLLPRIIRLALRLPDLIQSPVPLKHKHKNASLSLQSQIISCLLANAFLCTFPRNTLKK 270  
 QY 652 SEYSSYPDINFNRLPGRSSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--L 703  
 DB 271 SEYSTPDPINFNRLYQSTGPAVLEKLCIMHYFRRVCPPTERDASNNVPTGVVTFVRSGLP 330  
 QY 704 EDFPEWERCEKLL--TRLHVTVEGTIEGNGQGMLOVDFAHRFVGGVTSAGLVQVEIRFL 761  
 DB 331 EHLIDWSQAAPLGDVPLHVDVDAEGTIEGIGLLQVDFANKVLGGVGHGCVQVEIRFV 390  
 QY 762 INPELIVSRFLTEVLHDNECLITGTQEYSEYTGVAETVYRWARSHDRSERRDWRRTTE 821  
 DB 391 ICPELLVGLKFTCELRPFALVMLGAERYSNITGYAGSFEWSGNFEDSPTRDSSGRQTA 450  
 QY 822 IVALDALHFRYLDQVPEKIRRELKAYCGFLR-----PGVSSENLSAVATGNWCGGA 875  
 DB 451 IVALDALHFAQSHHQYREDLMERELNKAYIGFVHMVMTTPPG-----VATGNWCGGA 502  
 QY 876 FGDDARLKALIQILAAVAERDVVYTFGDSSELMDIYSMTHTFLTERKLTVCGEVYKLLLR 935



Db 503 FGSDSYLKALQMLVCAQLGRFLAYTTEGNEFRDDFHEMULLFRNDGTTVQQLWS-ILR 561  
Qy 936 YNNEECNCSPPG-----DIKLYPFIYHVESCTQTNTOPGORTGA 977  
Db 562 SYSRLIKESSKEPRENKASKKLYDFI---KEELKVRDVEGEGASA 606

RESULT 15

AAU76023  
ID AAU76023 standard; protein; 768 AA.  
XX AC AAU76023;  
XX  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE Fruit fly poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX  
XX Fruit fly; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
KW Huntington's disease; Parkinson's disease.  
XX  
OS Drosophila melanogaster.  
XX  
XX  
PN US6337202-B1.  
XX  
XX 08-JAN-2002.  
PD  
PF 23-FEB-2000; 2000US-00511477.  
XX  
XX 01-MAY-1998; 98US-0083769P.  
PR 30-APR-1999; 99US-00302812.  
XX  
XX (KENT) UNIV KENTUCKY RES FOUND.  
PA  
XX  
XX Jacobson WK, Jacobson EL, Ame J, Lin W;  
XX  
XX MPI: 2002-163240/21.  
DR N-PSDB; ABK14934.  
XX

Novel isolated and purified poly (ADP-ribose) glycohydrolase protein which catalyses release of ADP-ribose from ADP ribose polymer, useful for treating neoplastic and neurological disorders, heart attack and stroke.  
XX  
XX Example 13; Col 71-76; 81pp; English.  
XX

The present invention relates to a new poly(ADP-ribose) glycohydrolase (PARG) protein which catalyses release of ADP-ribose from an ADP (adenosine diphosphate)-ribose polymer. The PARG molecule of the invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy, reperfusion following ischaemia, heart attack, stroke, neurodegenerative diseases, neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases, CC and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs. Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the fruit fly PARG protein of the invention. This protein is one of several PARG proteins (AAU76020-AAU76024) of the invention

XX Sequence 768 AA;

Query Match 20.5%; Score 1063.5; DB 5; Length 768;  
Best Local Similarity 41.5%; Pred. No. 6.2e-77;  
Matches 244; Conservative 93; Mismatches 198; Indels 53; Gaps 15;

Qy 417 RVPKAEKKEKCEMKHQRTERKIPKIPPHL--SPDKKWLGTPEEM--RMRPGGRL 472

Db 45 RMSKSPGGGISEIETEETEE-----PENLANSLLDSWRGVSMEAHHRNQPFLENL 94  
Qy 473 PPLRPSANHTVTIRVDLLRIGEVKPPPTTHFKDLWDNKHVMPQSEONLYPVEDENGERA 532  
Db 95 PPVTAGNLHRVMYQ--LPITETP-PPYKSPGKWDSEHVRLLPCAPESKYPRENPDSGTT 150  
Qy 533 AGSRWELIQTALLNRLTRFQNLKDAILKYNVAVSKKWDFTALIDFWKVLBEAAQHLVQ 592  
Db 151 IDFRWEMIERALLQPIKTCEBELQAAIISYNTTYRDQWHERALHQLDEELDESETRVFFE 210  
Qy 593 SILPDMVKIALCLENICTOPTILKQKMHNSITWSQOIASILLANAFCTEPRRNA-KMK 651  
Db 211 DLLPRIIRLARLPDLIQSPVPLKHKHNASLSQQOISCLLANAFCTEPRNTLKRK 270  
Qy 652 SEYSYPDIINFNRLEGRSSRKPEKLKTLFCYFRV--TEKK-----PTGLVTFTRQS--L 703  
Db 271 SEYSTFPDINFNRLYQSTGPAVLEKLCIMHYFRVCPETERDASNVPTGVVTVFRSGLP 330  
Qy 704 EDFPWERCEKLL--TRLHVTYEGTIEGNGQGMLOVDFAFRFVGGVTSAGLVOEIRFL 761  
Db 331 EHLIDWSQSAAPLGDVPLHVDAGCTIIEDEGIGLLQVDPAFKYLGGLGGVGHGCVQVEIRFV 390  
Qy 762 INPELLIVSRLEFTEVLHDNECLIIITGTQYSEYTYAETYRMARSHEDRSERDDQRRTTE 821  
Db 391 ICPELLYVKLFTECLRPFEALVMLGAERYSNYTYAGSFWSGNFEDSTPRDSSGRRTA 450  
Qy 822 IVAIDALHFRYLDQFVPEKIRRELNKAYCGFLR-----PGVSEMLSAVATGNWCGGA 875  
Db 451 IVAIDALHFAQSHHQYREDLMERELNKAYIGFVHMVTPPEG-----VATGNWCGGA 502  
Qy 876 FGGDARLKALIQILAAVAERDVVFTFGDSSELARDIYSMHTFLTERKLTGVGVVKILLR 935  
Db 503 FGGDSYKALIQMLVCAQLGRFLAYTTEGNEFRDDFHEMULLFRNDGTTVQQLWS-ILR 561  
Qy 936 YNNEECNCSPPG-----DIKLYPFIYHVESCTQTNTOPGORTGA 977  
Db 562 SYSRLIKESSKEPRENKASKKLYDFI---KEELKVRDVEGEGASA 606

Search completed: May 26, 2004, 18:39:56  
Job time : 118.936 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 81.5151 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-2

Perfect score: 5184

Sequence: 1 MSAGPGCEPCTKRPRWDA...YHVESCTQTNPQORTGA 977

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5184	100.0	977	9	US-09-973-451-2
2	4635.5	89.4	976	9	US-09-973-451-4
3	4332.5	83.6	968	9	US-09-973-451-6
4	1063.5	20.5	768	9	US-09-973-451-8
5	594	11.5	546	12	US-10-425-114-60000
6	465	9.0	726	9	US-09-973-451-10
7	205.5	4.0	200	12	US-10-424-599-184988
8	184	3.5	180	12	US-10-424-599-156445
9	167	3.2	31	9	US-09-973-451-11
10	166	3.2	1462	14	US-10-287-218-17
11	164.5	3.2	300	12	US-10-424-599-233915
12	162.5	3.1	747	12	US-10-425-114-67803
13	162.5	3.1	1109	12	US-10-425-114-72939
14	162	3.1	748	9	US-09-864-761-43244
15	161	3.1	691	15	US-10-108-260A-4345

16	159	3.1	6642	15	US-10-369-493-5013	Sequence 5013, Ap
17	158.5	3.1	1973	16	US-10-197-824-34	Sequence 34, Appl
18	158.5	3.1	2047	12	US-10-346-863-2	Sequence 2, Appli
19	158.5	3.1	3863	16	US-10-197-824-7	Sequence 7, Appli
20	157.5	3.0	1163	12	US-10-221-278-197	Sequence 197, App
21	157.5	3.0	1163	15	US-10-291-172-197	Sequence 197, App
22	156	3.0	2665	9	US-09-864-761-34248	Sequence 34248, A
23	156	3.0	3664	12	US-10-263-929-143	Sequence 143, App
24	156	3.0	3664	14	US-10-177-293-423	Sequence 423, App
25	155.5	3.0	566	12	US-10-424-599-172054	Sequence 172054,
26	155.5	3.0	1501	9	US-09-924-154-17	Sequence 17, Appl
27	155.5	3.0	1804	14	US-10-275-140-2	Sequence 2, Appli
28	153	3.0	913	13	US-10-117-604-4	Sequence 4, Appli
29	152	2.9	914	15	US-10-369-493-1851	Sequence 1851, Ap
30	152	2.9	2697	15	US-10-144-198-12	Sequence 12, Appl
31	151.5	2.9	479	15	US-10-369-493-18078	Sequence 18078, A
32	151.5	2.9	665	10	US-09-820-843A-107	Sequence 107, App
33	149.5	2.9	1743	10	US-09-882-227-624	Sequence 624, App
34	149.5	2.9	1927	12	US-10-282-122A-58729	Sequence 58729, A
35	149.5	2.9	2224	14	US-10-115-563-14	Sequence 14, Appl
36	149	2.9	29	9	US-09-973-451-12	Sequence 12, Appl
37	148	2.9	3551	12	US-10-263-929-144	Sequence 144, App
38	147.5	2.8	1489	15	US-10-115-482-18	Sequence 18, Appl
39	147	2.8	26	9	US-09-973-451-19	Sequence 19, Appl
40	146.5	2.8	845	12	US-10-205-331-61	Sequence 61, Appl
41	146.5	2.8	1210	14	US-10-205-823-264	Sequence 264, App
42	146	2.8	1974	15	US-10-369-493-6395	Sequence 6395, Ap
43	145.5	2.8	447	10	US-09-794-422-44	Sequence 44, Appl
44	145.5	2.8	677	14	US-10-060-036-180	Sequence 180, App
45	145.5	2.8	677	14	US-10-060-036-4552	Sequence 4552, App

## ALIGNMENTS

### RESULT 1

US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF,  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match	100.0%	Score 5184;	DB 9;	Length 977;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 977;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	MSAGPGCEPCTKRPRWDAATSPPAASDARSPGQRRLVLSKDAPVQFRVPPSSGCA	60	
Db	1	MSAGPGCEPCTKRPRWDAATSPPAASDARSPGQRRLVLSKDAPVQFRVPPSSGCA	60	
Qy	61	GRAGCHRSATSLVFKQKTTISWMDTKGIKIVSESLSKNNNTRESMSSVKDNFY	120	

```

Db 61 GRAGQHRGSATSLVFKQKTTISWMDTKGKIVTESSELSHSENNNTREESMMSSVQKONFY 120
Qy 121 QHNNEKLENVSQLGFDKSPVKGQTOYLKQHTAAACKWQNGPHSHRLLSEPPAVTLVP 180
Db 121 QHNNEKLENVSQLGFDKSPVKGQTOYLKQHTAAACKWQNGPHSHRLLSEPPAVTLVP 180
Qy 181 EQFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKQTMEDQEGREARSHQKCGK 240
Db 181 EQFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKQTMEDQEGREARSHQKCGK 240
Qy 241 ACHPAEACAGCQOEEDTVVSEPLSDTSGEDVGTGLKNANRLNQESSLGNSPPFKESE 300
Db 241 ACHPAEACAGCQOEEDTVVSEPLSDTSGEDVGTGLKNANRLNQESSLGNSPPFKESE 300
Qy 301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRGSSA 360
Db 301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRGSSA 360
Qy 361 KGGEIRLHFQEGGESRAGMNDVNAKRPSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420
Db 361 KGGEIRLHFQEGGESRAGMNDVNAKRPSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420
Qy 421 AEDKREKQCECMKHORTERKIPKYPHLSPPDKMLGTPTEEMRMPCRGIRLPLRPSAN 480
Db 421 AEDKREKQCECMKHORTERKIPKYPHLSPPDKMLGTPTEEMRMPCRGIRLPLRPSAN 480
Qy 481 HTVTIRVDLLRIGEVKPPPTHEKDLMDNKHKVMPCEQNLVPVEDENGERAAGSRWELI 540
Db 481 HTVTIRVDLLRIGEVKPPPTHEKDLMDNKHKVMPCEQNLVPVEDENGERAAGSRWELI 540
Qy 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFMDKVLBAEAQHLYQSILPDMVK 600
Db 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFMDKVLBAEAQHLYQSILPDMVK 600
Qy 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTPPRRNAKMSSEYSSYPDI 660
Db 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTPPRRNAKMSSEYSSYPDI 660
Qy 661 NFNLFGESRRKPEKLTLCFYFRRVTEKKTGLVTFTRQSLDFPEWERCCKLLTRLH 720
Db 661 NFNLFGESRRKPEKLTLCFYFRRVTEKKTGLVTFTRQSLDFPEWERCCKLLTRLH 720
Qy 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRILFTEVLHDNE 780
Db 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRILFTEVLHDNE 780
Qy 781 CLIIITGTEQSEYTGAEYRWARSHEDSRDDQWRRTTEIIVADALHFRYLDQFVPE 840
Db 781 CLIIITGTEQSEYTGAEYRWARSHEDSRDDQWRRTTEIIVADALHFRYLDQFVPE 840
Qy 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWGCAGFGGDARLKALIQILAAVAERDVVY 900
Db 841 KIRRELKAYCGFLRPGVSSNLSAVATGNWGCAGFGGDARLKALIQILAAVAERDVVY 900
Qy 901 FTFGDSELMRDYSMHTFLTERKLTGVEYVKLLRYNNECRNSTPGPDIKLYPFIYHA 960
Db 901 FTFGDSELMRDYSMHTFLTERKLTGVEYVKLLRYNNECRNSTPGPDIKLYPFIYHA 960
Qy 961 VESCTQTNPQORTGA 977
Db 961 VESCTQTNPQORTGA 977

```

RESULT 2

```

US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US2002013228A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

```

```

; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
;
US-09-973-451-4

```

Query Match 89.4%; Score 4635.5; DB 9; Length 976;

Best Local Similarity 89.5%; Pred. No. 0;

Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

```

Qy 1 MSAGPGCEPCTKPRWDAAATSPPAASDARSFPGRQRRVLDSDAPVQFRVPPSSGAL 60
Db 1 MNAGPGCEPCTKATRWCAATTS-PAASDARSFPGRQRRVLDPKDAHVQFRVPPSSPACVP 59
Qy 61 GRAGQHRGSATSLVFKQKTTISWMDTKGKIVTESSELSHSENNNTREESMMSSVQKONFY 120
Db 60 GQAGQHRGSATSLVFKQKTTISWMDTKGKIVTESSELSHSENNNTREESMMSSVQKONFY 119
Qy 121 QHNNEKLENVSQLGFDKSPVKGQTOYLKQHTAAACKWQNGPHSHRLLSEPPAVTLVP 180
Db 120 QHNNEKLENVSQLGFDKSPVKGQTOYLKQHTAAACKWQNGPHSHRLLSEPPAVTLVP 179
Qy 181 EQFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKQTMEDQEGREARSHQKCGK 240
Db 180 EQFSNANVDQSPKDDHSDTNSERSDNOQFLTHVKLANAKQTMEDQEGREARSHQKCGK 239
Qy 241 ACHPAEACAGCQOEEDTVVSEPLSDTSGEDVGTGLKNANRLNQESSLGNSPPFKESE 300
Db 240 SCHPGEDCASCQDEIDVWPKSLSDVSGEDVGTGSKNDKLIQESCLGNSPPFEKSE 299
Qy 301 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRGSSA 360
Db 300 PESPMVDVNSKNSCODSEADSETSPGDEQEDSSAQTANKPSRFPQREADTELKRGSSA 359
Qy 361 KGGEIRLHFQEGGESRAGMNDVNAKRPSTSSLNVECRNSKQGRKDSKITDHFMRVPK 420
Db 360 KGGEIRLHFQEGGESRAGMNDVNAKRPSTSSLNVECRNSKQGRKDSKITDHFMRVPK 419
Qy 421 AEDKREKQCECMKHORTERKIPKYPHLSPPDKMLGTPTEEMRMPCRGIRLPLRPSAN 480
Db 420 AEDKREKQCECMKHORTERKIPKYPHLSPPDKMLGTPTEEMRMPCRGIRLPLRPSAN 479
Qy 481 HTVTIRVDLLRIGEVKPPPTHEKDLMDNKHKVMPCEQNLVPVEDENGERAAGSRWELI 540
Db 480 HTVTIRVDLLRIGEVKPPPTHEKDLMDNKHKVMPCEQNLVPVEDENGERAAGSRWELI 539
Qy 541 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFMDKVLBAEAQHLYQSILPDMVK 600
Db 540 QTALLNRLTRPQNLKDAILKYNVAYSKKWDFTALIDFMDKVLBAEAQHLYQSILPDMVK 599
Qy 601 IALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTPPRRNAKMSSEYSSYPDI 660
Db 600 IALCLPNICTQPIPLLKQKNHSITMSQEQIATSLANAFCTPPRRNAKMSSEYSSYPDI 659
Qy 661 NFNLFGESRRKPEKLTLCFYFRRVTEKKTGLVTFTRQSLDFPEWERCCKLLTRLH 720
Db 660 NFNLFGESRRKPEKLTLCFYFRRVTEKKTGLVTFTRQSLDFPEWERCCKLLTRLH 719
Qy 721 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRILFTEVLHDNE 780
Db 720 VTYEGTIEGNGQGLQVDFANFRVGGVTSAGLVQEEIRFLINPELIVSRILFTEVLHDNE 779

```

QY 781 CLIIITGEQSEYTGVAETVYRWARSHEDRSERDDWQRTTEIIVADALHFRYLDQFVPE 840  
Db 780 CLIIITGEQSEYTGVAETVYRWARSHEDRSERDDWQRTTEIIVADALHFRYLDQFVPE 839  
QY 841 KIRRELKAYCGFLRPGVSSSENISAVATGNWCGAGGADARLAKALQILAAAIAERDVVY 900  
Db 840 KIRRELKAYCGFLRPGVSSSENISAVATGNWCGAGGADARLAKALQILAAAIAERDVVY 899  
QY 901 FTFGDSELMRDYISMHTFLTERKLTIVGEVYKLLRYNNEECRCNSTPGDPDKLYPIYHA 960  
Db 900 FTFGDSELMRDYISMHTFLTERKLTIVGEVYKLLRYNNEECRCNSTPGDPDKLYPIYHA 959  
QY 961 VESCTQTTNPGQRTG 976  
Db 960 VESCAETADHSGQRTG 975

RESULT 3  
US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match 83.6%; Score 4332.5; DB 9; Length 968;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGFGCEPCTKRPWDAAATSPPAASDARSPPGQRRVLDSDKAPVQPRVPPSSGCGAL 60  
Db 1 MSAGFGCEPCTK-ARWGAAGTSAPTASDSRSPGQRRVLDSDKAPVQPRVPPSPACVS 59  
QY 61 GRAGGHRGATSILVFKQKTTISMTTKGKTVESSELSHSENNTRSESMSSVQKDNFY 120  
Db 60 GRAGGHRGATSILVFKQKTTISMTTKGKTVESSELSHSENNTRSESMSSVQKDNFY 116  
QY 121 QNMMEKLENVSLQGLDKSPVEKGTQVLKQHOQAAMCKWQNEGPHGSRLESPPAVTLVP 180  
Db 117 PHKVEKLENVQNLNDKSPTEKSSQYLNQQTASVCKWQNEGPHGSRLESPPAVTLVP 176  
QY 181 RQFSNANVDQSPKQDHSSTNSEEGRDNOQFTHVKLANAKVTMEDEQREARSHQCKG 240  
Db 177 KQLSNANIGQSPHTDHSSTNSEEGRDNOQFTHVKLANAKVTMEDEQREARSHQCKG 233  
QY 241 ACPAEACAGCQOETDVVSEPLSDTSGEDVGTGLKVNRLNROESSLGNSSPPEKSE 300  
Db 234 SRQSVADCTGCCQEEVDVLPESPLSDVGAEDTGTGPKNDKLTQGESLGDSPPEKSE 293  
QY 301 PESPMDVNSKNSCODSEADETSPGFQEQEDSSAQANKPSRFPQREADTELKRSQA 360  
Db 294 PESPMDVNSKNSCODSEADETSPGFQEQEDSSAQANKPSRFPQREADTELKRSQA 352

QY 361 KGGEIRLHFQFEGESRAGMNDVNAKRPGSTSLNVECRNSKHQGRKDSKITDHFMRVPK 420  
Db 353 KGSEVELHFFQB-GENNAGTSDLNKAPSGNSSSLNVECRSSKHQGRKDSKITDHFMRISK 411  
QY 421 AEDKREQCEQEMKHQRTERKIPKYPHLSPPDKWGLTPIEMMRMPCRGIRLPLRPSAN 480  
Db 412 SEDRRKEQCEVHRQRTERKIPKYPHLSPPDKWGLTPIEMMRMPCRGIRLPLRPSAS 471  
QY 481 HTVTIRVDLLRLLRGEVPEKPTTHFKDLWDNKHVMPCEQNLVPEVDENGSRAGSRWELI 540  
Db 472 HTVTIRVDLLRLLRGEVPEKPTTHFKDLWDNKHVMPCEQNLVPEVDENGSRAGSRWELI 531  
QY 541 QTALLNRLTRPQNLKDAILKYNVAYSKWDFDTALIDFMDKVLSEEAQHLVQSLPDMVK 600  
Db 532 QTALLNRLTRPQNLKDAILKYNVAYSKWDFDTALIDFMDKVLSEEAQHLVQSLPDMVK 591  
QY 601 IALCLPNICTQPIPLKQNMHSITMSQEQIASLLANAFCTFPPRNNAKMSSEYSPDI 660  
Db 592 IALCLPNICTQPIPLKQNMHSITMSQEQIASLLANAFCTFPPRNNAKMSSEYSPDI 651  
QY 661 NFNRLEFEGSRSSRKPEKLTFCYFRVRYTEKKTGLVTFTRQSLDEPPEWERCEKLTIRLH 720  
Db 652 NFNRLEFEGSRSSRKPEKLTFCYFRVRYTEKKTGLVTFTRQSLDEPPEWERCEKLTIRLH 711  
QY 721 VTYEGTIEGQGMQLQVDFANRFVGGVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 780  
Db 712 VTYEGTIEGQGMQLQVDFANRFVGGVTSAGLVQBEIRFLINPELIVSRFLTEVLDHNE 771  
QY 781 CLIIITGEQSEYTGVAETVYRWARSHEDRSERDDWQRTTEIIVADALHFRYLDQFVPE 840  
Db 772 CLIIITGEQSEYTGVAETVYRWARSHEDRSERDDWQRTTEIIVADALHFRYLDQFVPE 831  
QY 841 KIRRELKAYCGFLRPGVSSSENISAVATGNWCGAGGADARLAKALQILAAAIAERDVVY 900  
Db 832 KIRRELKAYCGFLRPGVSSSENISAVATGNWCGAGGADARLAKALQILAAAIAERDVVY 891  
QY 901 FTFGDSELMRDYISMHTFLTERKLTIVGEVYKLLRYNNEECRCNSTPGDPDKLYPIYHA 960  
Db 892 FTFGDSELMRDYISMHTFLTERKLTIVGEVYKLLRYNNEECRCNSTPGDPDKLYPIYHA 951  
QY 961 VESCTQTTNPGQRTG 976  
Db 952 VESCAETADHSGQRTG 975

## RESULT 4

US-09-973-451-8  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-973-451-8



```

Db 57 SRDVFVSSQSDSDQSDAENPEIAKEVSENCENLTETLKISNTIESLDNVTSEHTLDN 116
QY 292 ---SPPEKESESPMDV-----DNSKNSQDSEADEETSPGF--DEQE--- 331
Db 117 HKSTEPMEEDVNNKSNIDVAINDEDDDELVEENKEMRDGEQVQQLSQDLFADDQELIE 176
QY 332 -----DSSSQAQKPSFPQPREADTELKSSAKGGBIRLHFQPEGESR 377
Db 177 YGIMKDTTQLDITDSEVTAQQEMIEETEDS-----TFVGDSDK 219
QY 378 AGMNDVNAKPGSTSSLNVECRNSKHGRKDSKITDHFMRVPAKEDKRCQCEMKHQTE 437
Db 220 A-----TKTVTSSTSSFLST-----VSTCEAPAKGRARMYQKELE 254
QY 438 RKTPKYTPPHLS--PDKKWLGTPIEMRMRPCGIRLPPLRPSANHTVTIRVLLRIGEV 495
Db 255 KHVIAFTGEGNLTLPD-----LNKVDPRNY-----RYCTI 285
QY 496 PKPPP--THFKDLMDNKH---VKMPCSQNLVYVEDENGERAAGSRWELIQTALLNRLTRP 551
Db 286 PN--FPASQGLREDNRVGPXIVLPQRWREF-----DSRGR----- 320
QY 552 QNLKDAILKYN---VAYSKKWDFTA---LIDFWD---KVLSEAAQHLYSILPDM 598
Db 321 ---RDSYFFFKRLDGLVKCYKTYGPMFVGLLHNMWEPDPDITYKLPALEMYKEMSEL 377
QY 599 V-----KIALCLPNICTOPIPLLKQMHSTIIMSQEQIASLANAFAFFCTFPR 645
Db 378 VGREEVLEKFAVARIAKTAEDILPERIYRLGDV--ESATLSHKQCAALVARMEF----- 431
QY 646 RNAMKSESVSSPDINFNRLFEGRSSRKEKLTLCFYFRVTEKPTGLVTF--TRQSL 703
Db 432 ---ARPSPS-----FCRLSSDKICVEKFLFTFYDKMSMDPPDGAVSRLTKMDK 483
QY 704 EDPPEWERCEKLLRLHVTY--EGTIEGNGQMLQVDFAFRFVGGVTSAGLVQEEIRFL 761
Db 484 DTFNEEWKOKLRLSEVEFFEDWLEDTAL--CTQVDFANEHLGGVVLHNGSVQEEIRFL 542
QY 762 INPELIVSLRFTVELDHNELIITGTOYSEYTYGAYETRYWA-----RSHEDRSE--RDOV 815
Db 543 MCPENMVGMLCEKMKQLEAISIVGAUVFSYTYGYGHTLKWELQPNHNSRQNTNEFRDF 602
QY 816 QRTTEIIVALDALHPR-----YLDQVPEKIRRELKAYCGRLRGVSENLSAVATGN 870
Db 603 GRURVETIADLFLKSGKLDQTEQNLKANIIREMKKASIGFMSQGPFTNI--PIVTCW 661
QY 871 WCGAFGSGDARLKALIQILAAVAERDVVYTFGDSLMRDYISMHTELTERKLTVGE 928
Db 662 WCGAFNGDKPLKFLIQVIAAGVADRPFLHFCSEGEPELAACKCKIIRMKQKDVTLGX 719

```

```

RESULT 7
US-10-424-599-184988
; Sequence 184988, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184988
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(200)

```

```

; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138059C.1.pap
US-10-424-599-184988

Query Match 4.0%; Score 205.5; DB 12; Length 200;
Best Local Similarity 29.4%; Pred. No. 9.1e-08;
Matches 52; Conservative 26; Mismatches 70; Indels 29; Gaps 4;

QY 626 MSOEQIASLANAFAFFCTFPRRNAMKSEYSYP-----DINFNLEGRSSRKPDEK 676
Db 33 LTQETSAPPSRTFLCLFP-----VSDRPGIHLPMINSVSSGSLYEDYSQKPEST 83
QY 677 LKTLFCVFRVTEKKTGLVTFTRQSL-----DFPE---WERCEKLLTRLHVTYEG 725
Db 84 TGRAHYGQRISSEMPKGVISFERKVLFPKNDSIHISYPANFWSAISPLCFEVEHSSG 143
QY 726 TIEGNGQMLQVDFAFRFVGGVTSAGLVQEEIRFLINPELIVSLRFTVELDHNEL 782
Db 144 LIEDQSSGAELVDLANKLGGALGRGCVQEEIRFMVSPLEAGMLFLPAMADNEAI 200

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pap
US-10-424-599-156445

Query Match 3.5%; Score 184; DB 12; Length 180;
Best Local Similarity 29.4%; Pred. No. 3.7e-06;
Matches 48; Conservative 24; Mismatches 41; Indels 50; Gaps 4;

QY 845 ELNKAYCGFL-----RP----- 856
Db 7 EINKAFCGFLYQCKYQYQKILQENGTSALFYAATSTSMETDEGEISNHNKITSQNDYH 66
QY 857 GYSSNLSAVATGNWCGGAFGDARLKALIQILAAVAERD-VYTFGDSLMRDYISM 915
Db 67 GMDQGNITGVATGNWCGGAFGDPEVKTIQWLAAASQALRPFIAYTYTFG-LEALQSLDEV 125
QY 916 HTFLTERKLTVEGVYKLLRLYYNECRNCSTPGPDIKLYPRTY 958
Db 126 AHWILSQRTVQDGLWNLMTIYSINRSKGTENVGFLQWLPLPSIY 168

RESULT 9
US-09-973-451-11
; Sequence 11, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; ;
; TITLE OF INVENTION: (PARG) ENZYMES,

```

```

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE OF INVENTION: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-973-451-11

Query Match          3.2%; Score 167; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 5.7e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 771 LFTEVLDRNECLLIITGTSQYSEYTGATYR 801
Db 1 LFTEVLDRNECLLIITGTSQYSEYTGATYR 31

RESULT 10
US-10-287-218-17
; Sequence 17, Application US/10287218
; Publication No. US20030198975A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: AZIMZAI, Yalda; AU-YOUNG, Janice K.
; APPLICANT: BATRA, Sajeev; BAUGHN, Mariah R.
; APPLICANT: BECHA, Shanya D.; BOROWSKY, Mark L.
; APPLICANT: BUFORD, Neil; DING, Li
; APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.
; APPLICANT: GANDHI, Ameena R.; GIETZEN, Kimberly J.
; APPLICANT: GRIFFIN, Jennifer A.; HAPALIA, April J.A.
; APPLICANT: HONCHELL, Cynthia D.; LAL, Preeti G.
; APPLICANT: LEE, Soo Yeun; LU, Dying Aina M.
; APPLICANT: ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
; APPLICANT: REDDY, Roopa; SANJAYWALA, Madhu, M.
; APPLICANT: TANG, Y. Tom; WALIA, Narinder K.
; APPLICANT: WANG, Yu-mei, E.; WARREN, Bridget A.
; APPLICANT: XU, Yuming; YANG, Junming
; APPLICANT: YAO, Monique G.; YUR, Henry
; APPLICANT: ZERARUADIAN, Yeganeh
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PI-0417 USA
; CURRENT APPLICATION NUMBER: US/10/287,218
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: PCT/US02/11152
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US 60/349,705
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/295,263
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/295,340
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/293,727
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/291,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/291,662
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: US 60/287,228
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/286,820
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/283,294
; PRIOR FILING DATE: 2001-04-11
; Remaining Prior Application data removed - See File Wrapper or PALM.

```

```

; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 17
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030198975A1 5734806CD1
US-10-287-218-17

Query Match          3.2%; Score 166; DB 14; Length 1462;
Best Local Similarity 19.7%; Pred. No. 0.0022;
Matches 136; Conservative 97; Mismatches 253; Indels 204; Gaps 28;

QY 39 VLDSKQAPQVRVPSSSGC--ALGRAGQHRGSATSIVFKQKTTITSMMDTKITKIVTES 96
Db 606 VIPLSDSPPPSRMRKPGQKPKPSYERIQEMKAKTTHLLPIQSTY-SLANIK--ETGSSSS 662
QY 97 LHSKNNNTREESNMSSV-----QKDNFYQHNMEKLENV-SQLGFDKSPV 140
Db 663 YHKREKNSSESQSTYSKYSDRSSESSPRSRSRSSYSRSTRSLASSHSRSPS 722
QY 141 EKG---TQYLKQHQTAAAMCKW---QNEGPHSERLLESEPPAVTLVPEQFSNANV----- 188
Db 723 SRHSRNKYSDHSQCSRSSSYTSISSDDGRRAKRLRSSGKNSVSHKSSSEKTLHS 782
QY 189 -----DQSPKDDHSDTNS--EESRDNQPLTHVLKLANAKQTWEDEQGRARSHQCGK 240
Db 783 KYVKGDRDRSCVRKYSSRSLSLDYSSDSEQ--SSVQATQSAQKE-KQGOMERTHNQEK 839
QY 241 ACHPAEACACQOEETDVSESPLSD--TGSEDEVGTGLKXANRLNRQESSLGNSPPEKE 298
Db 840 -----NRGEKSKSERECPHSKKRTLENLSDHLRNGSKPKRK-----NYAGSKWD 885
QY 299 SEPEPMQV-DNSKNSQDSEADEETSFGDFDEQDSSAQTANKPSRFQPREADTELKRX 357
Db 886 SESNSRDVTKNKNDGSHPSDDKEEGEATSDSESEVSEIHKVKPT-----TK 933
QY 358 SSAKGGEIRLHFQEGGESAGMNDVN-----AKRPGSTSLNVECRNSK-----Q 403
Db 934 SS-----TNTSLPDDNGAWKSSKQRTSTSDSEGCSCNSNNRGPQKHK 977
QY 404 HGRKDSKITDHFMRVPAEDKREKQCEMKHQRTERKTPKIPPHLSPDKKWLGPPIEMR 463
Db 978 HGSKENLKRH---TKVKEKLGKKKKKAPKQAFHWQ-----1017
QY 464 RMPCRGIRLPLRPSANHTVIRVDLLRIGEVKPPPTTHFKDLMDNKHVKMPCSEQNLYP 523
Db 1018 -----PLEFGESEEEEDDKQVTKESKVKV-- 1043
QY 524 VEDENGERAAGRWELLQIOTALNRLTFPQNLKDAILKYNTA-----YSKKWDFTALIDFWD 579
Db 1044 --SENNE-----TIKONILKTEKSEEDLSGKHDTVTVSSDLD 1079
QY 580 KVLFEAEAAQHYQSILPDMVKIALCLPNICTQPIPLLKQNMHSITMSQEQIASLLANAF 639
Db 1080 QFTKDDSKLSISPTALNTENVA-CLQNI-----QHVEESVNGVEDVLTQDDNME 1129
QY 640 FCTFPRRNAAKQSEYSS-----YPDIN 661
Db 1130 ICT-PDRSSPAKVEETSPLGNARLDTPDIN 1158

RESULT 11
US-10-424-599-233915
; Sequence 233915, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

```

/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
/ FILE REFERENCE: 38-21(53223)B  
/ CURRENT APPLICATION NUMBER: US/10/424,599  
/ CURRENT FILING DATE: 2003-04-28  
/ NUMBER OF SEQ ID NOS: 285684  
/ SEQ ID NO 233915  
/ LENGTH: 300  
/ TYPE: PRT  
/ ORGANISM: Glycine max  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: PAT\_MRT1847\_5324C.1.pep  
US-10-424-599-233915

Query Match 3.2%; Score 164.5; DB 12; Length 300;  
Best Local Similarity 18.0%; Pred. No. 0.00027;  
Matches 53; Conservative 47; Mismatches 66; Indels 129; Gaps 5;  
QY 577 FWDVLEAPAHQYLSILPDMVKIALCLPNI-----CTQPIPLL 616  
Db 76 FPDVMSGEBSSKWFQEVLPALGNLLRLSPLESYHNTDNMAIDGAGAMLTALRL 135  
QY 617 KQWNSHITMSQIASILLANAFCTFPRNAXMKSEYSSYP-----DINFNRLF 667  
Db 136 DSQOPGIVFLITQELIALLSCLFLFP-----VSDRPVHILPMINFDVLFGLYD 186  
QY 668 GRSSRKEPKLTLCYFRRVTEKPTGLVTFTRQSLDFPWERCEKLLRLHVTYEGTI 727  
Db 187 DYSQKENKICVHVFORISSEWPKGIA----- 215  
QY 728 EGNQGMQVDFANRFVGGVTSAGLVQVEIRFLINPELIVSRKLTTEVLDPHNECLITGT 787  
Db 216 -----YNERIEIVGV 225  
QY 788 EYSEYTGVAETVWASHEDSRDDQWRITTEIIVADLHFRVYLDQVPPEKI 842  
Db 226 ERFSGYTDHASSPRFSRAKAGREEDPVGRRTKTDLS-----EKYFFPPKM 270

RESULT 12  
US-10-425-114-67803  
/ Sequence 67803, Application US/10425114  
/ Publication No. US20040034888A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Liu, Jingdong  
/ APPLICANT: Zhou, Yihua  
/ APPLICANT: Kovalic, David K.  
/ APPLICANT: Screen, Steven E.  
/ APPLICANT: Tabaka, Jack E.  
/ APPLICANT: Cao, Yongwei  
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
/ FILE REFERENCE: 38-21(5313)B  
/ CURRENT APPLICATION NUMBER: US/10/425,114  
/ CURRENT FILING DATE: 2003-04-28  
/ NUMBER OF SEQ ID NOS: 73128  
/ LENGTH: 747  
/ TYPE: PRT  
/ ORGANISM: Zea mays  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: LIB3689-225-C11\_FLI.pep  
US-10-425-114-67803

Query Match 3.1%; Score 162.5; DB 12; Length 747;  
Best Local Similarity 20.4%; Pred. No. 0.0015;  
Matches 101; Conservative 72; Mismatches 221; Indels 101; Gaps 15;  
QY 19 AATSPPAASDA-----RSPFGRRVLDSDAPVQFRVPRVSSGCGALGRAGQH 66  
Db 17 AAVSVEGSAKKGKNAEDETEKAVSAKKQKTVPEKAVPVKKQPPPKKVESSESDSS 76  
QY 67 RGSATSLVFKQKTTISMTDKGIKTVESSLHSEKNNNTREESM--MSSVQKDNFYQHNM 124

Db 77 DSEVEVKVQPKKVIQPKKSTQPAKQ\_ESSDDSSSESSDDPKKPVAGKKPPASSSSS 135  
QY 125 EKLENSQLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGPHSERLLESEPPAVTILVPEQFS 184  
Db 136 DSSSDSDSDSDEEPAKPTVLSKK-----PTVLSKKPTVLSKKPVTAV 179  
QY 185 NANVDQSSPKDHSPTNSSESRNQOPLTHVKLANAKQTMEDQGRREARSHQKCGKACHP 244  
Db 180 SNGSKQVKPDSSSSDSDEDEKPAASLKFPVVASVQKKTQESDSDSDSDEDESDIP 239  
QY 245 AEACAGCQOBTDWSESPLSDTSGEDVGTGLKNANRLNR-----QESSLGNSSP 294  
Db 240 AKAAVVAKKKESSSSSDSDSDESEDE---KPAANLKPPVPVSKKTQESDSDSDS 295  
QY 295 FEKESF--PESPMVNSK--NSCQDSEADDEETPGDEQDSSSAQTANKPSRFPQREAD 351  
Db 296 DEDESDDIIPAKAQQVAKKKESSGSSSDSDSE---FESDDSDNAKTV-----QPAKAA 345  
QY 352 TELRKSSAKGGEIRLHFQEGGESRAGMNDVNAKPGTSSLNVE----- 397  
Db 346 AIKKEDSSESE--TDSQSDSEDSPEKPTVPAKPLATNKKNEQSSSEKSDSDSDSD 403  
QY 398 -----CRNSKQHG-----RKDSKITDHFMRVPAKAEKQKQCEMKHOR-- 435  
Db 404 ESDDEPPQKTKNSAPSGAATAATNVSKKSSDEDESESSDDDEDTEQVETKTSVV 463  
QY 436 TERKIPKYPHPLSP 450  
Db 464 TTQK-----PAHNEP 473

## RESULT 13

US-10-425-114-72939  
/ Sequence 72939, Application US/10425114  
/ Publication No. US20040034888A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Liu, Jingdong  
/ APPLICANT: Zhou, Yihua  
/ APPLICANT: Kovalic, David K.  
/ APPLICANT: Screen, Steven E.  
/ APPLICANT: Tabaka, Jack E.  
/ APPLICANT: Cao, Yongwei  
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
/ FILE REFERENCE: 38-21(5313)B  
/ CURRENT APPLICATION NUMBER: US/10/425,114  
/ CURRENT FILING DATE: 2003-04-28  
/ NUMBER OF SEQ ID NOS: 73128  
/ LENGTH: 1109  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: Clone ID: LIB4119-145-F11\_FLI.pep  
US-10-425-114-72939

Query Match 3.1%; Score 162.5; DB 12; Length 1109;  
Best Local Similarity 18.8%; Pred. No. 0.0027;  
Matches 138; Conservative 111; Mismatches 308; Indels 177; Gaps 29;  
QY 76 KOKTITSWMDTKGIKTVESEL-HSKENN-----NTRSEMSVSVQKDNFYQHNEK-- 126  
Db 438 KRPLENLPTDQETVDINSVSEKKNIMITLTNIEHNLKSEKQKQKQFENKL 497  
QY 127 -----LNVSQLGFDKSPVEKGTQYLKQHQTAAAMCKWQNEGPHSERLLESEPPAV 176  
Db 498 IKSEIKDITLQTVLDSVQETGEKANIQAVDSE--VGLTKEDTQ-----EKLGEDDKTQK 551  
QY 177 TLVPE-----QFSNANVDQSPKDDHSDTNSSESRD-----NQOFLTH-----VKL 217  
Db 552 DVISNTSDVIGTYEADVAQKVDSEADTQSDNDGKVVVGVQKLINKPMVGPAGGTKE 611





Db 555 SPTDESIQKEGSG-----KALSAEMEYEMMHKTHKYKAPPAANERD 599  
QY 580 KYLEEAE-----AQHLYQSILPD 597  
Db 600 EVFEKEPLYGGMLEIDYIYESLVED 624  
RESULT 15  
US-10-108-260A-4345  
; Sequence 4345, Application US/10108260A  
; Publication No. US20040005560A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20040005560a1el full length cdna  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/10/108,260A  
; CURRENT FILING DATE: 2002-03-27  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4345  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-108-260A-4345  
Query Match 3.1%; Score 161; DB 15; Length 691;  
Best Local Similarity 19.6%; Pred. No. 0.0018;  
Matches 143; Conservative 101; Mismatches 291; Indels 196; Gaps 35;  
QY 41 DSKDAVQVRFPSPSSGSCAL---GRAGQHRGATSIVFKQKITITSMWDTKGKTVSEES- 96  
Db 31 DTSGSVARRAPKQKQSCILNVOSRGDTSGSVARRAPKQKQSS-----VVVIDSDSD 84  
QY 97 --LHSEKNN-----NTREES-----MMSSVQKQNFYQHNMEKLENVSQ 132  
Db 85 EECHTHEEKAKLEINSDDSEPECCHVKPAIQEPPIVISDDDDNDNDNGDLEVPDDNSD 144  
QY 133 LGFDKSPVEKGTQYLKQHQTAAMCKWQNEGPHSERLLESEPPAVTLVPEQFSNANVDQSS 192  
Db 145 -----DSEAP--DDNSDDSEAPDDNSDDSEAPDDNSDDSE 177  
QY 193 PKDDHS-----DTNSEESKRDNOQFLTHVKLANAKOTMEDEQGREARSHQKCGKACHPAE 246  
Db 178 APDDNSDDSDVPDDNSDDSD-----NSDSSDDNSDDSDVPDDKSDSDVPDD 227  
QY 247 ACAGCQOBTDVV-----SESP---LSDTGEDVTGLKNANRLNRQSSLGNSPPFEK 297  
Db 228 S-----SDSDVPDDSDSEAPDDSDSDSEAPDDSDSDSEAPDDSDSDSEAPDDSDSDS 282  
QY 298 ESEPESPMDVDNSKNSQDSEADETS PGFD---EOEDSSACTANKPSRFQPREADTEL 354  
Db 283 EASDDSDSDSEASDDSDSEAPDDKSDSDSDVPDDSDSDSDSDSDSDSDSDSDSDSDSD 342  
QY 355 RKRSSAKGGE-----IRLHFQEGEGESRAGMNDV---NAKEPGSTSSLNVECRNSK-- 402  
Db 343 NEGQIASDEEELVEAAAVSQH-----DSSDDAGEQLGENLSKPPSDPEANPEVSEKLP 398  
QY 403 -----QGRKDSKITDHFMRVPK-----AEDKRKEQCEMH-----QRT 436  
Db 399 TEEEPAPVVEQSGKRKSK-TKTIVPEPRKQTKTKNIVEPPPRKQTKTKNIVEPLRKA 457  
QY 437 ERKIPKYPH--LSPDKKWLGTPIEMR--RMPRC---GIRLPLRPSANHT----- 482  
Db 458 KTKNVSVTPGHKKGKSGKKGKGAKEKRTKTPCKVPCFCFLQDLEKSKKYSGKNLKN 517  
QY 483 ----VTIRVDLLRIGVPPFPFTHKDLMDNKHVKMP--CSEQNL-YPVEDENGERAAGS 535  
Db 518 KDELVQRIYDLFNRVCDKKLPEKLRIGWNNQVKTAGLCSTGEMWYP-----KWR 568  
QY 536 RWELIQTALLNRLTRPQNTKQAILKYNVAYSKKWDFALI-----DFWDKVLBEAEAOHLY 591  
Db 569 RFAKIQIG-LKVCDSADRIKDTLI-HEMCHAASWMLIDGTHSHGDANKYARKSNRIH-- 624

QY 592 QSILPDMVKIALCLPNICTQPIPLLKQKANHSHITMSOEQIA-----SLLANAFCTPFRN 647  
Db 625 ----PELPVTRC-HNY-----KINYKVHYECTGCKTRIGCVYTKSLDTSRFIC----- 667  
QY 648 AKMKGEYSYSP 658  
Db 668 AKCKGSLVMVP 678

Search completed: May 26, 2004, 19:18:52  
Job time : 84.5151 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.8764 Seconds  
(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-2  
Perfect score: 5184  
Sequence: 1 MSAGPGCEPCTKPRWDAAA.....YHAVESCTQTNPQGORTGA 977

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	522	10.1	997	B84726	probable poly (ADP-ribose)
2	465	9.0	726	T21138	hypothetical prote
3	331	6.4	364	A84726	probable poly (ADP-ribose)
4	185.5	3.6	1320	S57113	BUD4 protein - yea
5	175	3.4	1359	T34036	hypothetical prote
6	166	3.2	1403	A47328	natural killer cel
7	164	3.2	1115	T19137	hypothetical prote
8	162.5	3.1	1071	E85343	hypothetical prote
9	161.5	3.1	646	S15901	chromogranin B pre
10	159	3.1	1535	T39042	hypothetical prote
11	159	3.1	6642	T29757	protein UNC-89 - C
12	158.5	3.1	1576	T03277	pol protein - yeas
13	158	3.0	489	A45988	dentin matrix acid
14	157.5	3.0	1231	T18532	serine/threonine pr
15	157	3.0	1192	A71623	probable secreted
16	156.5	3.0	1200	A46194	neurofilament prot
17	155.5	3.0	522	C36608	hypothetical prote
18	155.5	3.0	1463	T30290	AAS surface protei
19	155.5	3.0	1804	T34518	nestin - golden ha
20	154.5	3.0	611	T06458	nucleolin homolog
21	154	3.0	406	S38170	SRP40 protein - ye
22	153	3.0	3122	T17202	DNA-directed DNA p
23	153	3.0	5327	T13564	microtubule-associ
24	152.5	2.9	915	B59433	chromosome 5 GAP-1
25	152	2.9	914	S48333	ORC1 protein - yea
26	152	2.9	1507	B47328	natural killer cel
27	151.5	2.9	665	B71609	hypothetical prote
28	151.5	2.9	1213	A58198	serine/proline-ric
29	151	2.9	1110	I51116	NF-180 - sea lamp

30	151	2.9	2248	2	A35938	profilaggrin - hum
31	150.5	2.9	532	2	T06029	hypothetical prote
32	150.5	2.9	1269	2	F84730	probable myosin he
33	150.5	2.9	1298	2	I54367	X-linked nuclear p
34	149.5	2.9	1927	2	G64585	cag pathogenicity
35	149.5	2.9	3147	2	T18674	hypothetical prote
36	148	2.9	635	2	T09648	nucleolin homolog
37	148	2.9	677	1	S09078	chromogranin B pre
38	148	2.9	3225	2	I52300	giantin - human
39	148	2.9	3259	1	A56539	giantin - human
40	147.5	2.8	913	2	T52485	neurofilament prot
41	146.5	2.8	845	2	A45669	neurofilament trip
42	146.5	2.8	1210	2	I39410	AP-4 protein, spli
43	146.5	2.8	1489	2	S60416	DNA helicase YGL15
44	146.5	2.8	1641	2	I38614	helicase II - huma
45	146.5	2.8	1791	2	T02345	hypothetical prote

## ALIGNMENTS

### RESULT 1

B84726  
probable poly (ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84726  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.P.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: B84726  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-997 <STO>  
A:Cross-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g31870  
A:Map position: 2

Query Match	10.1%	Score 522;	DB 2;	Length 997;
Best Local Similarity	32.0%	Pred. No. 6.3e-24;	Mismatches 116;	Indels 114;
Matches 136;	Conservative 59;			Gaps 12;
QY	578	WDKVLERAQAHLVQSILPDMVKIALCLPNT-----CTQPIPLKQKMHNS	623	
Db	79	FDELIDEXESKRWFDIIPALASLLQFSLLEVHFQADNIVSGIKTGLRLNLSQQAGI	138	
QY	624	ITMSQEQIASLLANAFCTFPRNAKMSVSSYPDINFR--LPEGSSSRKPEKLTILF	681	
Db	139	VFLSQELIGALLACSFCLFPDNRGAK----HLFVINFDHLSLYISYSQSESKIRCI	194	
QY	682	CYFRVRTEKPTGLVTFTRQ--SLEDPEWERCCKLRLHVTVEGTTEGNGQGMQVDF	739	
Db	195	HYFERFCSCVIGIVSFERKITAAPDAFNSKDVSL-----YQ-----PDNALEVDF	242	
QY	740	ANRFVGGVTSAGLVQBEIRFLINPELVSRFLTEVLDHNECLITGTPEQVSEYTGAE	799	
Db	243	ANKVLGGSLSRGCVQBEIRFPMINPELIGMLFLPRMDNEAIEIVGAERFSCYTGASS	302	
QY	800	YRWARSHEDRSERDDQRRTEIIVADAL-----HFRYLDQVPEKIRELANKVCGF	853	
Db	303	FRFAGEYIDKKAMPFKRRRTIIVADALCTPKMRFKDIC-----LLREINKALCGF	355	
QY	854	LR-----	855	
Db	356	LNCSEKANEHQNI FMEGDGNEIQLVNRGDSGLLRTETATSHRTPLNDVEMREKPNANLI	415	
QY	856	-----PGVSSGNL--SAVATGNWCGAFGGDARLKALIQILAAVAERDVV-YTFFGDSE	907	
Db	416	RDFVVEGVGDHEDHDDGVATGNWCGVFGGDPPELKATIQWLAASQTRRPFISYTFG-VE	474	



RESULT 4  
S57113  
BUD4 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein J1905; protein YUR092w  
C;Species: Saccharomyces cerevisiae  
C;Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Oct-1999  
C;Accession: S57113; S61952  
R;Ramezani Rad, M.; Kirchrath, L.; Hollenberg, C.P.  
submitted to the Protein Sequence Database, September 1995  
A;Reference number: S57111  
A;Accession: S57113  
A;Molecule type: DNA  
A;Residues: 1-1320 <RAM>  
A;Cross-references: EMBL:Z49592; MIPS:YUR092w  
R;Sanders, S.L.; Herskowitz, I.  
submitted to the EMBL Data Library, November 1995  
A;Reference number: S61952  
A;Accession: S61952  
A;Molecule type: DNA  
A;Residues: 'MHDAESTVDSLLKEIDNEMEQTKSNITQNGSEDTFHNWKLPLQHGDDTWMELVKNHNRNATENSRRSP', S  
202-212, 'D', 214-1007, 'A', 1009-1320 <SAN>  
A;Cross-references: EMBL:U41641; NID:g1151116; PIDN:AAB17116.1; PID:g1151117  
C;Genetics:  
A;Gene: SGD:BUD4  
A;Cross-references: MIPS:YUR092w; SGD:S0003852  
A;Map position: 10R

Query Match 3.6%; Score 185.5; DB 2; Length 1320;  
Best Local Similarity 16.9%; Pred. No. 0.0031;  
Matches 185; Conservative 156; Mismatches 361; Indels 391; Gaps 45;

QY 70 ATSLVFKQKTIISWMTK-----GIKVESLSLHSEN-----NNTRESMSSVQKDNFY 120  
DB 290 STSTKSESYIADYKVTROEDWDTKLHORESHANEQPAIIPOKDSSESTFTELNNESEF 349

QY 121 QHNM-----EKLNVSLQGD-----KSPVEKGTQYLKQ 149  
DB 350 QRFKDGEBYRIVOHRESLYQRTKSPENINGSIGVDHGAAVNEPLAKTS--AEE 407

QY 150 HQTAAMCKWQ--NEGPHSERLLE-----SEPPAVTLVPE- 181  
DB 408 HDLSSCEDQSQSEARNKORIEKEVETKDIETEKDESEYHKVENESEPEHPVLLPPL 467

QY 182 -----QFNSANVDQSPKDDHSD-----TNSEESRDNOQFLT 213  
DB 468 PRWEEIQFNEPFDIDENDTSDNDSILTRSMKPSYISIMHIQEEIKSNPSPESTANSQFSQ 527

QY 214 HVKLANAKQTMEDEQREARSHQ-----KCGKACHPAEACAGCQ--QEETDVVSESP-- 263  
DB 528 QSSITTA-STVDSKDKNGSTSFKEKPRIVRSRSHLYNPKSRVSLNYDNEYDILNSERN 586

QY 264 -----LSDTGEDVGTGLKNA-----NRLNRESSLGNSSPPPEKSESPSPMD 306  
DB 587 ALDPMERNLTLKRIQDNIRTOKGHAPLIRPSIMKLNGEDSGFN-----HFELEVPQOE 641

QY 307 VNKNKSCODSEADETSQFQEOEDSSAQATANKSRFPQREADTFLKRSAA----- 360  
DB 642 HENIPULSTHSEODITNVLGDEQKLPNTQD-----EATISIRIESAGDITFN 691

QY 361 KGGEIRLHFQFEGESRAGANDV-----NAKRPSTSSINVECRNSKO- 403  
DB 692 RGLLSLSFDEELGQDFANFLDALDHDSTSFNHGPDSSQFQDSSKFSNLSWESSYEL 751

QY 404 -----HGKDKSKITDHPMRVPKAEKKQCEMKHQTER-----KIP----- 441  
DB 752 KPPPSIRKQPIAPDVLOKL--LESPTKDDADLEKIREERITEPRTGLIGLMLKTPVKDVS 809

QY 442 -----KVIPP-----HLSPDKKWLK 456  
DB 810 IALAASIKYEAFSTDSRPEGMNNSDAITLNMFDFFEDKMTPTPTVRSISPIKRHVS 869

QY 457 TP-----IEEMRMPCRGIR--LPPIRPSANHTVTRVDLLRI 492

DB 870 SPFKVVKAGNKQNNENINIKABEEIEPMTQOETDGLKQDIPPLLAQTKONVAKETITQ 929  
QY 493 GEVPKFPPTHFXD-----LWDNKHVROMPCSEONLYPVEDENGERAAGSRW 537  
DB 930 LSEPDQVEQEFPMGPTLYLSIKAI STLALVGTK-----SHRATYAI VFDNGENVVQTPW 983  
QY 538 ELI-----QTALLNRLFPQNLKDAILKYNVAYS-K-KWDF TALID 576  
DB 984 ESLPYDGNIRINKFELP IDFKGTSTSSASSERDSYKCVITLCKYKPRHELVEIVD 1043  
QY 577 -----FWDKVLBEAEAHLYQSILPDMVKIALCLPNICTQPIPLLKQXN 621  
DB 1044 KYPVGKSTFGTKYKYEKKYVQKPKQDEWDYLFAQDGSFARC----- 1086  
QY 622 HGITMSQEQIASLLANAFCTPPRRNAKMKSESYSD--INENRLPFGSSRSKPKLKT 679  
DB 1087 -EIEINEE-----FLKNVAENT-SHMHNINKWSRIADKIHGSKRLYE-LPRKAPHKVAS 1139  
QY 680 L-----FCYFRVTEKPTGLVTTRQSLDDPP-EMERCEKLLTFL-----HVTVEGTIEGNG 731  
DB 1140 LDVEACFLERTS-----AFEQFPKQFSLVNVKIVSKYKLOQNIYREGYLLQDG 1186  
QY 732 QGMQLQVDFANRFV--GGGVTSGAGLVQEEIRFLINPELIV-----SRLFTE 774  
DB 1187 -GDLGKIENRPFKLHGSQSLSGVHEISRKAKIDINLLKVTKVLRNEDIQADNGGQNFDT 1245  
QY 775 VLDHNECLITGTQYSEYTVGAETVWARSHEDSERDDQWRRTTEIIVADALH---FR 831  
DB 1246 WVLNECFQLVFD-----GERITFNACSNEEKS---DWYNKLGQVVELNVFHPVWK 1296  
QY 832 RYLDQVPEKIR 844  
DB 1297 KYCEKLABEEKTR 1309

RESULT 5  
T34036  
hypothetical protein B0041.7 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T34036  
R;Fulton, R.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A;Description: The sequence of C. elegans cosmid B0041.  
A;Reference number: Z21466  
A;Accession: T34036  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1359 <FUL>  
A;Cross-references: EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CESP:B0041.7  
A;Experimental source: strain Bristol N2; clone B0041  
C;Genetics:  
A;Gene: CESP:B0041.7  
A;Map position: 1  
A;Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 3.4%; Score 175; DB 2; Length 1359;  
Best Local Similarity 18.5%; Pred. No. 0.014;  
Matches 141; Conservative 124; Mismatches 284; Indels 214; Gaps 32;

QY 87 KGIKTVESLSHSEKNNNTRESMSSVQKDNFYOHNMKLENVSLQGFDPKSPVEKGTQY 146  
DB 81 KSRKRAKES-ESDESDEEDRKSKSKKVKDQKKKSKKKKRTTSSSEDESDEREQ- 138

QY 147 LKQHOTAAAMCKNQNEGPHSERLLESEPPAVTLVPEQSNANVDQSPKDDHSDTNSSESR 206  
DB 139 -KSKKSKKTKKQTSSSESE--ESEEERKVKSKKKNKSKSVKRAETSESEDEDEKPSK 194

QY 207 DNQQFLTHVLANAKQTWEDEQGRARSHQKCGKACHPAEACAGCQOEETDVVSSPLSD 266  
DB 195 KSKKGLK--KKAKESESESEDEKVKSKKSKKVKVKESESEDEAPEKKTETKRRKSK 252

QY 267 TGSEVDGTGLKANRLNQESSLGNSSPPPEKSESE-----ESPMVDVNSKNQSDSEADEE 322

```

Db 253 TSSE- - - - -SSESEKDEEEESKSPKPKKKPLAVKLLSSDESEEDVE 300
Qy 323 TSPG- - - - -PDEQDSSAQTANKPSRFPREADTELKRSKSAKGGSIHLHFOFEG 373
Db 301 VLPQKKRGAVTLISDSEDEKQKSESEASVVEKVKSKAKAKQESSE- - - - -SG 350
Qy 374 GESRAGMNDVNAKPGSTSSLANVCRNSKHG- - - - -RKDSKITDHFMRVPKAEDEKEQCEM 431
Db 351 SDSSEGSITVNRK- - - - -SKKKEPKKKKGIIMDSKKQKQETIDAEAEKERRKLEK 404
Qy 432 KHQR- - - - -TERKIPKYPHLSPPDKKWLGTPIEENRMRPRCGIR 471
Db 405 KQKEFNGVILEEGEDLTEMLTGTSSQKRLKSV- - - - -LDPO- - - - -SSTVDEESKXP- - - - -452
Qy 472 LPLRPSANHTVTRVLLRIGEVKPPPTH- - - - -FKDLWNKHVKMPCSEONLYPVEDENGE 530
Db 453 - - - - -VEVHNSLVRI- - - - -LKPHQAHGIOFYD- - - - -CAPESLDRLDTEG- - - - -489
Qy 531 RAAGSRWELIQTALLNRLTRPQNLKDALIKYNVAYSKKWDFALIDFWDKVLEAEQAHL 590
Db 490 - - - - -SGGILAHCMGLGKT- - - - -LQVITFLHTVL- - - - -M 515
Qy 591 YQSILPDMVKIALCLP- - - - -NICTQPIPLLKOKMHSITMSQEQIASLLANAFCTPPRRNAK 649
Db 516 HEKIGEKCKRVLVVVKVNIINWF- - - - -KBFQKVLVNDDELDIDVN- - - - -559
Qy 650 MKSYSSYPDINFNR- - - - -LPEGRSRKPEKLTLCYFRVVT- - - - -EKPTGLVVTTRQ 701
Db 560 - - - - -ELDSYKTIEDRRALKAWHSSKTPSVMIIGYDLFRILTVEDDPKKKPKPNRRRLEK 616
Qy 702 SLEDF- - - - -PEWERCE- - - - -KLLTRLHVYEGT- - - - -LEGN- - - - -730
Db 617 AKEDFRKYLQNPQGMVWVCDBAHLKNDSDALSCKMVKILTRKICITGTPLONNLMEYH 676
Qy 731 - - - - -QCGML- - - - -QVDFANFVGGVTSAGLVQV- - - - -EIRFL 761
Db 677 CMNVFVKPGLLGTETEFANRV- - - - -NIINRGTKDASPLEVSFM 717

RESULT 6
A47328
natural killer cell tumor-recognition protein - human
N;Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C;Accession: A47328
R;Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A;Title: A cyclophilin-related protein involved in the function of natural killer cells.
A;Reference number: A47328; MUID:93133824; PMID:8421688
A;Accession: A47328
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1403 <AND>
A;Cross-references: NID:g181251; PIDN:AAA35734.1; PID:g181252
A;Experimental source: NK killer cells from adult blood
A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIPI:122800)
C;Genetics:
A;Gene: GDB:NKTR
A;Cross-references: GDB:137171; OMIM:161565
A;Map position: 3p23-3p21
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C;Keywords: alternative splicing; lymphocyte
F;60-230/Domain: cyclophilin homology <CYP>

Query Match 3.2%; Score 166; DB 1; Length 1403;
Best Local Similarity 19.7%; Pred. No. 0.053;
Matches 136; Conservative 97; Mismatches 253; Indels 204; Gaps 28;

Qy 39 VLSDKAPVFRVPPSSSGC- - -ALGRAGQHRGTSLSLVFKQKTTSMWDTKGIKTVSESS 96
Db 661 VIPLSDSPPPSRWPKGPKWPKYSYQEMKAKTTHLLPIQSTY-SLANIK- - -ETGSSS 717

```

```

Qy 97 LHSKENNTREESNMSSV- - - - -OKONFYQHNMKELENV-SQLGDFDKSPV 140
Db 718 YHKREKNSDQSTYKYSRDRSSSRSSRSYTRSRSLASSHSRSPS 777
Qy 141 EKG- - - - -TOYLKQHQTAAACKW- - - - -ONEGPHSERLLESEPPAVTLVPEQFNAV- - - - -188
Db 778 SRSHRNKYSDHQSCSSSYTSISSDDGRRAKRLSSGKKNSVSHKKHSSSEKILHS 837
Qy 189 - - - - -DOSSPKDDHSDTNS- - - - -EESRDNQOFLTHVKLANAKQTMEDBQGHREARSHQCGK 240
Db 838 KYVKGRRDRSCVKYSESRSLLDYSSDSEQ- - - - -SSVQATQSAQKE- - - - -KQGQWERTHNQEK 894
Qy 241 ACPAECACACQOEETDVSESLSD- - - - -TGSDEVGTGLKNAIRLNRRSESLGNSPPPEKE 298
Db 895 - - - - -NRGEKSKSRECEPHSKRTILKENLSDLRNGSKPKK- - - - -NYAGSKWD 940
Qy 299 SEPESPMDV- - - - -DNSKNSCQDSEADBEETSPGDEQDSSSAQTANKPSRFQPREADTELKR 357
Db 941 SESENREDVTYKNKNDSPSDKEEGEATSDSESEVSEIHKVKPT- - - - -TK 988
Qy 358 SSAGGSEIRLHFOFEGGESRAGMNDVN- - - - -AKRPGSTSLNVCEKNSK- - - - -Q 403
Db 989 SS- - - - -TNTSLPDDNGAWKSSKQRTSTSDSEGSNSNNRGKPKQHK 1032
Qy 404 HGRKDSKITDHFMRVPKAEDEKREQCEMKQRTKIPKYPHLSPPDKWLGTPIEEMR 463
Db 1033 HGSKENILKREH- - - - -TKVKREKLKGGKKGKPKKQAFHWQ- - - - -1072
Qy 464 RMPRCGIRLPLRPSANHTVTRVDLLRIGEVKPPFTPHFKDLWNKHVKMPCSEONLYP 523
Db 1073 - - - - -PLEGEEEEIEIDDKQVTSKEKVV- - - - -1098
Qy 524 VEDENGRAAGSRWELIQTALLNRLTRPQNLKDALIKYNVA- - - - -YSKKWDFALIDFWD 579
Db 1099 - - - - -SENNE- - - - -TIKONILKTEKSEEDLSGKHDTVTVSSDLD 1134
Qy 580 KVLBEAEACHLYQSILPDMVKIALCLPNICTQPIPLLKOKMHSITMSQEQIASLLANAF 639
Db 1135 QFTKDDSKJISPTALNTENVA- - - - -CIQNI- - - - -OHVESVPNGVEDVLQTDNDME 1184
Qy 640 FCTFPRRNKMKSEYSS- - - - -YPDIN 661
Db 1185 ICT-PDRSSPAKVEETSPGLNARLDTPDIN 1213

RESULT 7
T19137
hypothetical protein C09G1.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T19137
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19079
A;Accession: T19137
A;Status: preliminary; translated from GS/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1115 <WIL>
A;Cross-references: EMBL:Z50176; PIDN:CAA90539.1; GSPDB:GN00028; CESP:C09G1.2
A;Experimental source: clone C09G1
C;Genetics:
A;Gene: CESP:C09G1.2
A;Map position: X
A;Introns: 14/1; 49/2; 1095/2

Query Match 3.2%; Score 164; DB 2; Length 1115;
Best Local Similarity 19.5%; Pred. No. 0.051;
Matches 167; Conservative 98; Mismatches 306; Indels 286; Gaps 39;

Qy 1 MSGAGPCEP- - -CYKRRPRDAAATSPPAADARSFPQRQRRLVDSKADPQVFRVPPSSSGC 58
Db 106 ISKSTWNEPISCSKIKLSQKQTSPTQNDPDYFTAKPTDKITR- - - - -PSISRG 157

```

C;Genetics:  
A;Gene: AT9g29440  
A;Map position: 4

Query Match            3.1%; Score 162.5; DB 2; Length 1071;  
Best Local Similarity 21.5%; Pred. No. 0.06;  
Matches 111; Conservative 76; Mismatches 194; Indels 135; Gaps 24;

	Qy	42	SKDAPVQRVPSPSSGGCALGRAGORHGSATSLVKKTTTSWMDTKGIKVVESSLHSKE	101
	Dd	389	SRENSLENR--PNDFAKIGRKQPFG-----MDDINLSSADVLNKKOSSRASHSPPS	440
	Qy	102	NNNTREEGMMMS---SVQQDNFVQHNMEXLENVSQLGFCDKSPV-----	140
	Dd	441	SNFSDDNVDTALDHIDTDSPIFEKNFKQTIVGDRES-----YDSPVVVAAPAFDYSSFF	495
	Qy	141	-----EKGTY-----LKOHQAAMCKWNQG-----PHSERL	168
	Dd	496	KDPFDTEDAYHDPEOGLGFLSGSSSTSDHMPTEISSWSLEGHKDLGKLSSASTSQV	555
	Qy	169	LESPPAVTLPEQPSN-----ANVDQSP--KDHSHTNSESRNQQLTHVKLANA	220
	Dd	556	LEKEKPS---SPPTPDGFTSPASLUHEPESAKFYDRDSEEDNLGRLS--GRAEG	610
	Qy	221	KOTMEDEGREARSHQKCGKACHPA----EACAGCQOEETDVVSESPLSDTGSE--DYGT	274
	Dd	611	XSKLTAAQSHPMSEGPDLDLGRIYFPSTDTEQDGDDSKTQEESD--AETP--TGLKFGPLAS	665
	Qy	275	GKMANRLNRQESSILGNAPPPEKESPE-----SPMDVDNKSNCQDSEADEETSFGFDEQ	330
	Dd	666	GLENETTL-----PSYGSSPPRDKTSKSKEYLPTEVDPSPRSSLSLOTASSIRNELIYTQ	721
	Qy	331	EDSSSAQTANKPSRFOP-----READTELKRSSAKGEIRLHFPOEGGESRAGMN--D	382
	Dd	722	KASNDK---RPSSIPPPSSSDDESMLPKRVSFYQEKRTESRTRPTHLSGVSHKD	778
	Qy	383	VNAKRGPGSTSINVCRNSKHGRKDXKITDFHMVRYPKAEDKKEQCCEMKHQRT----	436
	Dd	779	LEEIPTRASTRSQRRTHKTPASASASYFTMS-SDDEDEXEVHRDTAHIQTRYISI	837
	Qy	437	-----ER-----KIPLY-----IPHLSPDKK	453
	Dd	838	SRRTKQGRRRSULTAKIDKXSVFDEESPPLKSPEAK	873

RESULT 9  
SI5901  
chromogranin B precursor - bovine  
C;Species: Bos primigenius taurus (cattle)  
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999  
C;Accession: SI5901; S20727; A61076; S21773; S39369  
R;Bauer, J.W.; Fischer-Colbraie, R.  
Biochim. Biophys. Acta 1089, 124-126, 1991  
A>Title: Primary structure of bovine chromogranin B deduced from cDNA sequence.  
A;Reference number: SI5901; PMID:91223091; PMID:2025642  
A;Accession: SI5901  
A:Molecule type: mRNA  
A;Residues: 1-646 <BI>  
A;Cross-references: EMBL:X55489; NID:g11; PIDN:CXX38846.1; PID:g12  
R;Grandy, D.K.; Leduc, R.; Makam, H.; Flanagan, T.; Dilberto, E.J.; Thomas, G.; Civelli,  
submitted to the EMBL Data Library, October 1990  
A>Description: Nucleotide and deduced amino acid sequence of bovine adrenal medulla Chromo-  
granin B protein.  
A;Reference number: S20727  
A;Accession: S20727  
A>Status: preliminary  
A:Molecule type: mRNA  
A;Residues: 21-63 'N', 65-92 'FRS', 96 'RA', 99-100 'S', 101-480 'L', 482-646 <GRA>  
A;Cross-references: EMBL:X55489; NID:g201; PIDN:CXX39109.1; PID:g202  
R;Flanagan, T.; Taylor, L.; Poulter, L.; Viveros, O.H.; Dilberto Jr., E.J.  
Cell. Mol. Neurobiol. 10, 507-523, 1990  
A>Title: A novel 1745-Dalton pyroglutamylyl peptide derived from chromogranin B is in the  
A;Reference number: A61076; PMID:91223542; PMID:1982622  
A;Accession: A61076

A:Molecule type: protein  
 A:Residues: 567-580 <FLA>  
 R:Dillen, L.; Boel, S.; de Potter, W.P.; Claeys, M.  
 Biochim. Biophys. Acta 1120, 105-112, 1992  
 A:Title: Mass spectrometric characterization of bovine chromaffin granule peptides related to the protein tyrosine phosphatase SH-PTPase  
 A:Reference number: S21773; PMID:92207983; PMID:1554736  
 A:Accession: S21773  
 A:Molecule type: protein  
 A:Residues: 567-580 <DIL>  
 R:Yoo, S.H.  
 Biochim. Biophys. Acta 1179, 239-246, 1993  
 A:Title: pH-dependent binding of chromogranin B and secretory vesicle matrix proteins to the protein tyrosine phosphatase SH-PTPase  
 A:Reference number: S39369; PMID:94032431; PMID:8218367  
 A:Accession: S39369  
 A:Molecule type: protein  
 A:Residues: 239-244; 562-565 <YOO>  
 C:Superfamily: chromogranin B precursor  
 C:Keywords: adrenal gland; glycoprotein; pituitary; pyroglutamic acid; sulfoprotein  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-646/Product: chromogranin B #status predicted <MAT>  
 F:567-580/Product: pyroglutaminyl peptide BAM-1745 #status experimental <MAT>  
 F:567-580/Product: pyroglutaminyl peptide BAM-1745 #status experimental <MAT>  
 F:36-57/Disulfide bonds: #status predicted  
 F:188/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:315/Binding site: sulfate (Tyr) (covalent) #status predicted  
 F:567/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 Query Match 3.1%; Score 161.5; DB 1; Length 646;  
 Best Local Similarity 20.7%; Pred. No. 0.035;  
 Matches 158; Conservative 87; Mismatches 249; Indels 269; Gaps 39;  
 15 RWDAAATSPPAASDARSPPGRRRLVDS--KDAPYQFRVPPSSSCALGRAGQHRGSATS 72  
 80 RFEVLLRDPA--DISEAPGLSSRE--DSGEGDAQPTVADTESGGHSRERAGEPPGSQVA 136  
 73 LVFKQKTTISWMDTKGKIVTSELSHSENNTRRESMSSVQKDNFVQHNMEKLENVSQ 132  
 137 -----KEAKTRYKSEKQNRREEMVK-----YQ-KRERGEVGE 169  
 133 LGFDKSPVEKTOYLKQHTAAMCKWQNEGHSERLLESEPPAVTLVPEQSNANVDQSS 192  
 170 ERLSEGPKQAQAFINQ-----RNQTPAKKEELVSR-----YDTQSARGLEKSH 213  
 193 PKDHSNTSESRDNDQOFLTHVLKANAKQTMEDQREARSHQKCGKACHPACAGCQ 252  
 214 SRERSQSEGETKQENW-----PQELQRHPEGEA-----PGES----- 249  
 253 QEETDVVSEPLSDTGSDDVGTGLKANRLNRQESSLGNPPFEKESPEPMDVDNSKN 312  
 250 --EEDA--SPEVD-----KRHSRPHHGRSRPDRSSQEGNPPL----- 284  
 313 SCQDSEADETSPGFDEQEDSSAQTANKPRQFQPREADTLKRSSAKGGEIRLHFQFE 372  
 285 -----EEESHVG-----TGNSEDEKARHPAFRALEGAAY-----GHEVRHSA-- 324  
 373 GGESRAGMNDVNAKPGSTSLNVECRNSKQHRKXDSKITDHFMRVPAKADKRQCEBMK 432  
 325 -----AQAPGDLQCARFGRCGRGHE-----QALRRPSEES--LEQENKR 361  
 433 HQTERKIPKIPPHLSDDKWLGTPIEM--RRMPRCGIRLPPLR-----PSANHTVIR 486  
 362 H-----GLSPDLNMAQGYSESEBERGPAFG-----PSYRARGGAAAYSTLIGQ 405  
 487 VDLRL-IGEVKPPFFTHFK--DLWDNKHVKMPCSEONLVPVEDENGRAAGRWELIQT 542  
 406 TDEKRFLE-----THRVQESQDKARRRLPGLRLNVDYGEKGEAARGKWQ----- 455  
 543 ALLNRLTRPQ-NLKDAILKYNVAYSKWDFTALIDFWDKVLEEAQHLQSYILPDWVKI 601  
 456 -----PQGDPRDA-----DENREEARLGR----- 474  
 602 ALCLNICTQPIPLKQKNHSHITMSQEQIASLLANAF-----CTFPPR----- 646  
 475 -----KQYAPHHIT--EKRLGELL--NPFYDPSQWKSRRFERKDPMDDSFL 516

QY 647 -----NAKMKSEYSSYPDINFN-----RLFE-----GRSSRKPEKLTLCFYFRRVTEKKP 692  
 DB 517 EGEENGGLTNEKKNFFFEYNDWWEKKPFEDVNWGTEKNPVPKLDLKROYDRVAELDQ 576  
 QY 693 TGLVTFTRQSLDEPWEKCEKLLTRLHVTYEGTIEGNGQGL 735  
 DB 577 --LLHYRKSAE--FPDFVDSQEQSPQH--TAENEERKAGQGV 615  
 RESULT 10  
 T49042  
 hypothetical protein TSP19.60 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
 C:Accession: T49042  
 R:Benes, V.; Wurmbach, E.; Drzonek, H.; Ansorge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.; submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z25014  
 A:Accession: T49042  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1535 <BEN>  
 A:Cross-references: EMBL:AL163972; GSPDB:GN00061; ATSP:TSP19.60  
 A:Experimental source: cultivar Columbia; BAC clone TSP19  
 C:Genetics:  
 A:Gene: ATSP:TSP19.60  
 A:Map position: 3  
 A:Introns: 38/3; 96/1  
 Query Match 3.1%; Score 159; DB 2; Length 1535;  
 Best Local Similarity 18.9%; Pred. No. 0.16;  
 Matches 182; Conservative 133; Mismatches 338; Indels 312; Gaps 47;  
 6 GCECTKTRPW-----DAAATSPPA--ASDARSPPGRRRLVDSKDAPV-----QFRVPP 53  
 DB 89 GCDSTLHAKRWEPPDNDHTNTIPEALLSSQNRSLSAEVESPEDGSRTPMRTTHREYNRP 148  
 QY 54 SSSGCGALGRAGHRSATSILVFQKTI--TSWMDTKG--LKTVESELSHSENN--NTR 106  
 DB 149 STS-----VERGYHPETVYKPTSDIRREWRRTDDFSETGSDVFTSERSPYNT 200  
 QY 107 EESMW-----SSVQKDNFYQHMEKL 127  
 DB 201 SNAQAQWAGHEGYADPPRVFPYPASPSSAYEYGYSSPPHGSYVASEQSYHQPNQF 260  
 QY 128 ENVSQLG-FDKSPVEKGTQY-----LKHQHTAAMCKWQNEGHE-- 164  
 DB 261 EQYREGMFOQSSVASPTFRPGTSDGKYHRSQSQQLHDLYHNLYEPSPR--SETPHHS 318  
 QY 165 --SERLLESEPPAVTLVPEQFSNANVDQSSPKDDHSDTNSSES--RDNQOFL----- 212  
 DB 319 VYSR--SYVPAAPHRSTYSEHSVGISK-----SDTSSEKSLRKNKRYVRERNPVVK 370  
 QY 213 THV-----KLANAKOTME-----DEQREARSHQKCGKACHPACAGCQ-----QE 254  
 DB 371 RHILPSAGGAPFATCYCLELLQLPQVSPQKQRYQVRCQ-----SCSGVLKFSIRE 423  
 QY 255 ETDVVSPLSDTGSDDVGTGLKANRLNRQESSLGNPPFEKESPEPMDVDNSKN 314  
 DB 424 KADTVLDSPP-----SFVDYGMDFADFTVTNHQDSA-----SEGHEINPDGSHLTC 469  
 QY 315 QDSADETSPGFDEQEDSSAQTANKP--SRFQPREADTLKRSS-----AKGGEIRLH 368  
 DB 470 LDDSGETIC-----KSDAVILSKILLETFKEDKIDIRNISTKFLDLKLEALQPH 522  
 QY 369 FQFEGSGRAGMNDVNAKPGSTSLNVECRNSKQHRKXDSKITDHFMRVPAKADKRQCE 428  
 DB 523 LKPMGNRLSEQPTSETTIGETSKIHL--QSQEAHSEKSTEMDNNDICDIEPE----- 575  
 QY 429 CEMKHQTERKIPKIPPHLSDDKWLGTPIEM--RRMPRCGIRLPPLRPSANHTVIRVD 488  
 DB 576 ----YKNEIKPEIIVGRSGES--LEVFPVQNERM----- 606



QY 489 LLRIGEVPPKPPPTFFKDLWDNKHVKMPCSEQNLYPVVEDEN-GERAAGSRWELIQTALLNR 547  
Db 607 -----SESS-----EEDERVAERSVSHSEELLYKKEYVS 635  
QY 548 LTRPQ--NLKDAILKYNVAYSKKWDFTALIDFWKVLBEAQAHLVQSI--LPD--MWKI 601  
Db 636 ETRVQSENTGD-----TYFEWEST-----QTEEYEGESTLEETEEIGDKPQLHLDS 681  
QY 602 ALCLPNICTQPIPLKQKMNHSITMSQEOIASILANAFCTPPRRNAKMSYSSYPDIN 661  
Db 682 RYTRDSICDSSLSYSGFFEDTSVKEDTEHISNTISDA-----KSATHTSPINSL--VN 732  
QY 662 FNRLFEGRSRKPEKLTILFCYFRRVTEKKPTGLVFTFRQSLDFEWEBCR----- 713  
Db 733 ENQRFIGHLDSGEATVTI-----SRQBE-----HSKYEYENLSERQELEMTISDRVT 780  
QY 714 -KLLTRLHVTYGTIGNGQGMQLQVDPANRFVGGVTSAGLVQBEIRFLINDELIVSRLF 772  
Db 781 CKLESRYETNE-TLEPSKVG-----BDGLSVSPFKVSEAFETIIEGKTGERLV 829  
QY 773 TEVLGH-NECLIIT-GTEQ--YSEYTGAEYRWARSHEDRDDQRRTEIIVAILDAL 828  
Db 830 SHQMSPNVENSYGLTLPVLEESGFAGERTWEGTIEDRA-----GL 873  
QY 829 HFRRY 833  
Db 874 HLEXY 878

RESULT 11  
T29757  
protein UNC-89 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 03-Dec-1999  
C:Accession: T29757  
R:Du, Z.; Le, T.T.; Wilson, R.  
submitted to the EMBL Data Library, May 1997  
A:Description: The sequence of C. elegans cosmid C09D1.  
A:Reference number: Z20679  
A:Accession: T29757  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-6642 <DUZ>  
A:Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89  
A:Experimental source: strain Bristol N2; clone C09D1  
C:Genetics:  
A:Gene: CESP:unc-89  
A:Map position: 1  
A:Introns: 1/72; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1;  
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 3.1%; Score 159; DB 2; Length 6642;  
Best Local Similarity 20.2%; Pred. No. 1.2; Mismatches 387; Indels 206; Gaps 45;  
Matches 185; Conservative 136; Mismatches 387; Indels 206; Gaps 45;

QY 9 PCTKRPRWDAATSPPAASDARSFPQGRVRLDSKDPVQFRVPPSSGCGALGRAGRHG 68  
Db 1407 PREKSPASTKKEKSPAABEVKS-PTKKEK---SPSSPTKKEKSPSPPTKKTGDEVKEKS 1462  
QY 69 SATSLVFKOKTITSMWDTK-----GKTVSESE-SLHSSKNNNTRE-----ESM 110  
Db 1463 PPKSPTKKEKSPKPEKDVKSPPVKKEKSPDATNIVEVSSETTIEKTETTTTETTTETHESEES 1522  
QY 111 MSSVQKDNFQHNWEXLENVSQLGPKSPVEKTKTVYLKQHTAAMCKWQNEGPHSERLLE 170  
Db 1523 RTSVKKKEKTEKPEKDEKPSPTKK--DKSPKSPITEIKSPVK-----KEKSPKVEEKP 1574  
QY 171 SEPPAVTLVPEQFSNANVDSSPKDHSIDTNSBESRDNQOFLTHVKLANAKTME-DEQG 229  
Db 1575 ASPTKEKSPK--PASPTKKESENEVKSPTKKEKSP--KSVVEELASPKKEKSPKADDDX 1631  
QY 230 REARSHQKCGKACHPAACAGCQOQET-DVWSSPLSDTSGSEVDGT-----GLKNNRNL 283

Db 1632 KSPTKKEKSPKSAVEDVKSPTKKEKSPKVEKEKTSPTKKESSPTKKTDDDEVKSPTKKE 1691  
QY 284 QRESSLGNSP--PPEKSEPESPMDVNSKNSQDSSEADEETSPPGDEQDSSSAOTANK 341  
Db 1692 KSPQVVEKPPASPTKKEKSP-----KSVVEEVKSPEKSPKAEKPKSPPTKKEKS 1743  
QY 342 PSRFQPREADTBLRK-RSSAKGGEIRLHFQFGESRAGMNDVNAKRP-----GSTSSLNV 396  
Db 1744 PEKSAAEVKSPTKKEKSPKSAEAKPKSPTKKSSPVKMAADDEVKSPTKKEKSPKVEE 1803  
QY 397 ECRNSKHQGRKDSKITDHFMRVP-----KAEDKRKEQC-EMHQRTERILPKY 443  
Db 1804 KPASTTKKEKTEKSAABEELKSPTKKEKSPSPPTKKTGDESEKSPKPEKSPKSPPTKK 1863  
QY 444 IPPHLSDDKWLGTPIEEMRRMPCGIRLPPURPSANHTVTIRVLLRLRIGVEPKPPPTHF 503  
Db 1864 SPFG-SPKKKKKSP--EAKKP-----APKLRDLKLTQVNTKTDLAHF-EVVVEHATEC 1914  
QY 504 KDLWONKHVKMPCSBQNLYPVDENGE-----RAAGSRWELIOTA 543  
Db 1915 KWFLDGKEI---TTAQGVTVSKDDQFEPKCSIDTTMFGSGTVSVVASNAAGS----VETK 1967  
QY 544 L-LNRLTRPONLKDALIKYNVAYSKKWDT-ALIDF-----WDKVEEAQAQ--LYQ--- 592  
Db 1968 TELKVLETPKE-----TKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKNYQ 2015  
QY 593 -----SILPDMVKIALCLPNICTQPIPLKQKMNHSITMSQEOIASILANAF 640  
Db 2016 NGNLLDGKNGVTIKNEENKSLIIPN--AQSGKITVEASNEVG-SSSSAQLTWNPPS 2072  
QY 641 CT-----PPR-----RNAKMSYSSYP-----DINFNRLFEGRSRKPEKLTFLC 682  
Db 2073 TPIVVDGPKSVTIKETETAEPKATISGPAATVKWTIN-EKIVE--ESRTITTIKTEDV 2129  
QY 683 YPRRVTEKK--PTGLVTFTRQSL--EDPEWE-----RCEKLLTRL--HYTEG---- 725  
Db 2130 YTLKISNAKIEQTGTVKVTAQNSAGDSKQADLKVEPNVKAPFKSKQLTDKVADEGEPLR 2189  
QY 726 ---TIEGNGQGMQLQVDPANRFVGGVTSAGLVQBEIRFLINDELIVSRILFTEVLHDNE-C 781  
Db 2190 WNELEIDGSPG-----TEVSWLLNGQPLTKSTVQVVDHGDGT 2227  
QY 782 LIITGTEQYSEYTG 795  
Db 2228 YHVTIAEAKPEMSG 2241

RESULT 12  
T03277  
pol protein - yeast (Candida albicans) retrotransposon pCal (fragment)  
C:Species: Candida albicans  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 29-Oct-1999  
C:Accession: T03277  
R:Matthews, G.D.; Goodwin, T.J.D.; Butler, M.I.; Berryman, T.A.; Poulter, R.T.M.  
J. Bacteriol. 179, 7118-7128, 1997  
A:Title: pCal, a highly unusual Ty1/copia retrotransposon from the pathogenic yeast Cand  
A:Reference number: Z14877; MUID:98037512; PMID:9371461  
A:Accession: T03277  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1576 <MAT>  
A:Cross-references: EMBL:AF007776; NID:g2636718; PIDN:AAC49878.1; PID:g2636719  
C:Genetics:  
A:Mobile element: retrotransposon pCal

Query Match 3.1%; Score 158.5; DB 2; Length 1576;  
Best Local Similarity 19.2%; Pred. No. 0.18;  
Matches 145; Conservative 103; Mismatches 267; Indels 241; Gaps 36;

QY 5 PCCEPCTKPRWDAATSPPAASDARSFPQGRVRL--DSKDAP-VQFRVPPSSSGCALG 61  
Db 660 PGDNFVSK-----PPQLGTETSVIGKSPKPTKPTKAPSIQGRDHRKSESTA-- 707





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 26, 2004, 17:20:49 ; Search time 15.6533 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-2

Perfect score: 5184

Sequence: 1 MSAGPGCEPCTKRPRDAAA.....YHVESCTQTTPQORTGA 977

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	180.5	3.5	1447	BUD4 YEAST	P47136 saccharomyc
2	174	3.4	1359	ATRX CAEL	Q9u7e0 caenorhabdi
3	166	3.2	1462	NKCR HUMAN	P30414 homo sapien
4	162	3.1	5147	PCLO HUMAN	Q9y6v0 homo sapien
5	161.5	3.1	646	SG1 BOVIN	P23389 bos taurus
6	159.5	3.1	513	DMPI HUMAN	Q13316 homo sapien
7	159	3.1	6832	UN89 CAEL	Q01761 caenorhabdi
8	158.5	3.1	5596	MDNI HUMAN	Q9nu22 homo sapien
9	156.5	3.0	773	ELAI MOUSE	Q8cb77 mus musculu
10	156	3.0	3644	MINT MOUSE	Q62504 mus musculu
11	156	3.0	3664	MINT HUMAN	Q96t58 homo sapien
12	155	3.0	489	DMPI RAT	P98193 rattus norv
13	154	3.0	406	SR40 YEAST	P32583 saccharomyc
14	153	3.0	913	IF38 HUMAN	Q99613 homo sapien
15	153	3.0	3122	DPOZ MOUSE	Q61493 mus musculu
16	152.5	2.9	915	CE05 HUMAN	Q9nyf5 homo sapien
17	152	2.9	914	ORC1 YEAST	P54784 saccharomyc
18	152	2.9	1453	NKCR MOUSE	P30415 mus musculu
19	149	2.9	503	DMPI MOUSE	O55188 mus musculu
20	149	2.9	845	NFM RAT	P12839 rattus norv
21	149	2.9	5038	PCLO MOUSE	Q9qvx7 mus musculu
22	148	2.9	633	MLH TETH	P40631 tetrahymena
23	148	2.9	677	SG1 MOUSE	P16014 mus musculu
24	148	2.9	3259	GOB1 HUMAN	Q14789 homo sapien
25	147	2.8	2004	MYG3 HUMAN	Q92794 homo sapien
26	146.5	2.8	1210	AF4 HUMAN	P51825 homo sapien
27	146.5	2.8	1489	YGF0 YEAST	P53115 saccharomyc
28	145.5	2.8	500	GAR2 SCHPO	P41891 schizosacch
29	145.5	2.8	677	SG1 HUMAN	P05060 homo sapien
30	145.5	2.8	2224	PAS HUMAN	P12259 homo sapien
31	145	2.8	950	STL2 MOUSE	Q99n50 mus musculu
32	144.5	2.8	668	SG1 FIG	Q9gl94 sus scrofa
33	144.5	2.8	757	CC5 SCHPO	P39964 schizosacch

34	144.5	2.8	1478	1	THO2 HUMAN	Q8ni27 homo sapien
35	144	2.8	598	1	CYL1 HUMAN	P35663 homo sapien
36	144	2.8	678	1	GRAP_PLAEP	P13816 plasmodium
37	144	2.8	1362	1	BUD4 HUMAN	Q60885 homo sapien
38	144	2.8	2459	1	MAPB RAT	P15205 rattus norv
39	144	2.8	5560	1	SPEN DROME	Q8sx83 drosophila
40	143.5	2.8	699	1	NPL4 HUMAN	Q14978 homo sapien
41	143.5	2.8	1147	1	CARB HUMAN	Q9bx17 homo sapien
42	143.5	2.8	1338	1	ACIN MOUSE	Q9jix8 mus musculu
43	143.5	2.8	5120	1	PCLO_CHICK	Q9pu36 gallus gall
44	143	2.8	2492	1	ATRX_HUMAN	P46100 homo sapien
45	143	2.8	2492	1	ATRX_PONPY	Q7yqm3 pongo pygma

## ALIGNMENTS

RESULT 1  
BUD4 YEAST STANDARD; PRT; 1447 AA.  
ID P47136;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Bud site selection protein BUD4.  
GN BUD4 OR YJR092W OR J1905.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96295555; PubMed=8707826;  
RA Sanders S.L., Herskowitz I.;  
RT "The BUD4 protein of yeast, required for axial budding, is localized to the mother/BUD neck in a cell cycle-dependent manner.";  
RL J. Cell Biol. 134:413-427(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ramezani Rad M., Kirchraht L., Hollenberg C.P.;  
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
CC -!- FUNCTION: Co-assembles with BUD3 at bud sites. BUD4 and BUD3 may cooperate to recognize a spatial landmark (the neck filaments) during mitosis and they subsequently become a landmark for establishing the axial budding pattern in G1.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- CAUTION: Ref.2 sequence differs from that shown in the N-terminal section due to frameshifts.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----

EMBL; U41641; AAB17116.1; -;  
EMBL; 249592; CAAB9620.1; ALT\_FRAME.  
EMBL; 249591; CAAB9619.1; -;  
PIR; S57113; S57113.  
GermOnline; 141924; -;  
SGD; S0003852; BUD4.  
InterPro; IPR001849; PH.  
Pfam; PF00169; PH; 1.  
SMART; SM00233; PH; 1.  
PROSITE; PS00003; PH\_DOMAIN; 1.  
ATP-binding. 181 189 ASP/GLU-RICH (ACIDIC).  
NP BIND 1175 1182 ATP (POTENTIAL).  
DOMAIN 1302 1413 PH.  
CONFLICT 291 328 FTASVEDLNISTNLFAGDSSQNNPVTADALIENDVV ->  
ANSLCGGFKHFNESRVLIPAKIIQSLMRMLLTMLC

```
FT CONFLICT 340 340 (IN REF. 2).
FT SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;
SQ SEQUENCE 1447 AA; 164485 MW; 1D056B9FF1B7067B CRC64;

Query Match 3.5%; Score 180.5; DB 1; Length 1447;
Best Local Similarity 16.8%; Pred. No. 0.0099;
Matches 184; Conservative 156; Mismatches 362; Indels 391; Gaps 45;

QY 70 ATSLVFKQKITSMWTK-----GIKIVBSLSHSEN-----NNTREBSMSSVQKDNFY 120
DB 417 STTKSESYIADYKVTROEDWDTKLHQSEHANEQPAIIPQKDSSEETFTLNNESEF 476
QY 121 QHNM-----EKLNVSQLGFD-----KSPVEKGTQVLKQ 149
DB 477 QRFKDGEEYRIVHESLYCQRTKSPENINGSEIGVDHGEAEVNEPLAKTS--ABE 534
QY 150 HQTAAMCKWQ--NEGPHSERLLE-----SEPPAVTLVPE-- 181
DB 535 HDLSSCEDQSVSEARKDRIEKEVETKDNIEKDESEYHKVEENEPEPHVPLPPL 594
QY 182 -----QFSNANVDQSPKODHSP-----TNSSESDNQOFLT 213
DB 595 PRWEEIOFNEPFDENTSDSIDLTKMPKSPDVISWHIQEIEKNSPESIANSPSQ 654
QY 214 HVLANAKQWEDQGREARSHQ-----KCGKACHPAEACAGCQ--QEETDVVSESP-- 263
DB 655 QSSITTA-STVDSKKGSTSFKPKPRIVSRSLYNPKSRVSSLYNYDNEYIILNSEWN 713
QY 264 -----LSDTGSBDVGTGLKNA-----NRLNRQSSSLGNSPPPKESPEPSMD 306
DB 714 ALDPMRENTLISKRIQDNIRTKGHAPLIRPSIMKNGEDSGFON-----HFLVEVQPE 768
QY 307 VDSKNSQDSEADEEISPGFDEQEDSSAQANKPSRFQPREADTELKXESSA----- 360
DB 769 HENIPLTHUSEQDITNVGLDEQKLPNTQD-----EASIRIETESAGDITFN 818
QY 361 KGGEIRLHFQEGGESAGMNDV-----NAKRGSTSSLNVECRNSKQ-- 403
DB 819 RGLDLSLSFDEELQDFANFLDALDHDSTSNHGPDDSSQFQDSSKKSFNLSWESSYEL 878
QY 404 -----HGRKDSKITDFMRVPAKAEKREKQCKMKHQRTER-----KTP----- 441
DB 879 KPPIRSIRKQIAPDVLQKL--LESDTKDDADLEKIRBERITEPTRTGLGIGMLKTPFVDVS 936
QY 442 -----KYPYP-----HLSPDKKWLQ 456
DB 937 IALAASIKGYEASPSDTSRPEGMNNSDAITLNNFDDPEEDKMTPTPVSISIPKRVHS 996
QY 457 TP-----IEEMRRMPRCGIR--LPPLRPSANHTVTIRVDLLRI 492
DB 997 SPFKVVKAGNKQENNEINIKAEIEEPTQETDGLAQDIPPLAQTKDNVAKEEITQ 1056
QY 493 GEVPKPPTTHPKD-----LMDNKHVKMPCSEONLYPVDENGERAAAGSRW 537
DB 1057 LEEPQDVEQEPFDMGTLYLSIKALSTLALYTK-----SHRATYAIYFDNGENVVQTPW 1110
QY 538 ELI-----QTALLNRLTRPQNLKDALTKUNVAYSK--KWFETALID 576
DB 1111 ESLPYDGNIRINKGEFELIDPKGKAFTSSASSERDSYKCVITLCKYKPRHBLVEIVD 1170
QY 577 -----FMDKVLAEAEAAQHYLSIPLDMVKIALCLPNICTOPTIPLKQMN 621
DB 1171 KVPVKGFFGKTKYKFKYKQKPKQDWDYLAQDGSFARC----- 1213
QY 622 HSITWSQBIASLILANAFCTFPFRNNAKMESESYSDP--INFNRLFEGSSRKPPEKLT 679
DB 1214 -EIEINEE-----FLKNVAFNT--SHMVMNINKWRIADKIHGSKRLYE--LPRKAPHKVAS 1266
QY 680 L---PCVFRVTEKKTGLVFTTQSLDFFP--EWERCERKLLTRL-----HVYETGTGNG 731
DB 1267 LDVACFLERTS-----AFQFPKQFSLNVKIVSKYKLOQNIYKQGLIQDG 1313
QY 732 QGMLQVDPANRFV---GGGVTSAGLVQBEIRFLINPELIV-----SRLFTE 774
```

```
DB 1314 -GDLKGKIENRFFKLGSQLSGYHESRKAIDINLLKVTKLVLNEDIQADNGQORFTD 1372
QY 775 VLHNDECLITGTQYSEYTYGAEYRWAHSHEDRSERDDMQRTTETIVDAIH---FR 831
DB 1373 WVLNFECFQLVFD-----GERITFNAECSEBKS---DWYINKLQVVELNVFHPQWVK 1423
QY 832 RYLDQFVPEKIRR 844
DB 1424 KYCKLAABEETR 1436

RESULT 2
ATRX CAEL STANDARD; PRT; 1359 AA.
AC QUTEO; Q02061;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
GN XNP-1 OR B0041.7.
OS Caenorhabditis elegans.
OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99365296; PubMed=10433961;
RA Villard L., Fontes M., Ewbank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RT the human XNP/ATR-X gene.";
RL Gene 236:13-19 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton R., Wohlmann P.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: Could be a global transcriptional regulator. Modifies
CC gene expression by affecting chromatin (Potential).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF134186; AAC55361.1; -
CC EMBL; AF000196; AAC24256.1; -
CC PIR; T34036; T34036.
CC WormPep; B0041.7; CE17314.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001650; Helicase C.
CC InterPro; IPR000330; SNF2 N.
CC Pfam; PF00271; Helicase C_1.
CC Pfam; PF00176; SNF2_N; 1.
CC SMART; SM00487; DEAD_C; 1.
CC SMART; SM00490; HELIC_C; 1.
CC DNA repair; PS00690; DEAD ATP HELICASE; FALSE NEG.
CC DNA-binding.
CC NP_BIND 496 503 ATP (POTENTIAL).
CC SITE 636 639 DEAD BOX.
CC DOMAIN 67 70 POLY-ASP.
CC DOMAIN 266 272 POLY-GLU.
CC FT DOMAIN 276 281 POLY-LYS.
CC FT DOMAIN 372 375 POLY-LYS.
CC FT DOMAIN 603 608 POLY-LYS.
CC FT DOMAIN 859 862 POLY-LYS.
CC FT CONFLICT 479 479 C -> F (IN REF. 2).
```

SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;  
 Query Match 3.4%; Score 174; DB 1; Length 1359;  
 Best Local Similarity 18.5%; Pred. No. 0.022;  
 Matches 141; Conservative 124; Mismatches 284; Indels 214; Gaps 32;  
 QY 87 KGITVESESLHSKNNNTREEMSSVQKQNFYQHNMEKLENVSLQGLPKFVKGITQY 146  
 DB 81 KSKRAKSES-ESDEDEEDRKKSKSKKKVQKKKKKRTTSSSEDESDEREQ- 138  
 QY 147 LKHQTAAMCKWQNEGPHERLLEPPAVTVLPQFNSNANTVQSPKDDHSDTSESR 206  
 DB 139 -KSKKKSKTKQTSESSE---ESEERKVKSKKQKSKVKGKGAETSESEDEKPSK 194  
 QY 207 DNQQLTHVLANAKQWDEQGRARSHQKCGKACHPAEACAGCQOQETDVVSEPLSD 266  
 DB 195 KSKGLK--KKAASRSESESEDEKVKSKKKKKVVKKESESEDEAPEKKTERRKRSK 252  
 QY 267 TGSEVDGTGLKNANRLNROESSIGNSPPPEKSESEP-----ESPMVDNNSKNSQDSEADE 322  
 DB 253 TSSEE-----SSESEKSDSEEEKSSPKPKKKKPLAVKLLSSDESESDVE 300  
 QY 323 TSPG-----FDEQDESSAQTANKSRFPQPREADTELKRSKAGGEIRLHFQEG 373  
 DB 301 VLPQKKKRGAVTLISDESEDKQKSEASDVEEKVSKKAKKQESSE-----SG 350  
 QY 374 GBSRAGMNDVNAKPGSTSSLVNCEKNSKHQ--RKDSKITDHFMRVPAEDKRKEQCEM 431  
 DB 351 SDSSEGSITVNRK-----SKXKEPEKCKKGIIMDSKLOKETIDAEAEKERRKLEK 404  
 QY 432 KHQR-----TERKIPKVIPEHLSPDKKWLGTPIEEMRMPCRGIR 471  
 DB 405 KQKEFNGIVLEGEDLTMLTGTSSQKRLKSVV---LDEP-----SSTVDEESKCP- 452  
 QY 472 LPPLRPSANHTVIRVDLRIIGEVKPPFPTH-FKOLMDNKHVMPCSEONLYPVDENGE 530  
 DB 453 -----VEVHNSLVRI--LKPHQAHGQFMVD-----CACESLDRLDTEG- 489  
 QY 531 RAAGSERWELIQTALLNRLTRPNQKXDAILKYNVASKKWDFTALIDFWDKVLAEAEQHL 590  
 DB 490 -----SGGILAHCMGLGKT---LQVITFLHTVL-----M 515  
 QY 591 YOSILPDMVKIALCLP-NICTOPIPLKQNMHSITMSQEQIASLANAFAFFCTFPRNAK 649  
 DB 516 HEKIGCKRVLVVPKVINWF---KEFQKWLVDNDEELDTIDV- 559  
 QY 650 MKSEYSSPDINFR--LPEGRSSRKPEKLTLCYFRVVT-----EKPTGLVTFTRQ 701  
 DB 560 ---ELDSYNTIEDRRALKAWHSSKTPSVMIITGYDLFRILTVEDDDPKKKKPNRRNRLEK 616  
 QY 702 SLEDF-----PWERCE-----KLLTRLHYTCT-LEGN----- 730  
 DB 617 AKEDFRKYLQNGPDMVQDDEAHKLNDDSAKCMVKILTRICLTGTPLQNNLMYEH 676  
 QY 731 -----GQGLM--QVDFANRFVGGVTSAGLVQE-----EIRFL 761  
 DB 677 CMVNFVKPGLLTKTEFANRFV--NIINRGRTKASPLEVSEFM 717

RESULT 3

NKCR\_HUMAN  
 ID NKCR\_HUMAN STANDARD; PRT; 1462 AA.  
 AC P30414;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE NK-tumor recognition protein (Natural-killer cells cyclophilin-  
 DE related protein) (NK-TR protein).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;

[1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE-Blood;  
 RX MEDLINE=93133824; PubMed=8421588;  
 RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,  
 RT Ortaldo J.R.;  
 RT "A cyclophilin-related protein involved in the function of natural  
 RT killer cells."; Proc. Natl. Acad. Sci. U.S.A. 90:542-546 (1993).  
 RN [2]  
 RP REVISIONS.  
 RA Anderson S.K.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 CC !- FUNCTION: Component of a putative tumor-recognition complex.  
 CC Involved in the function of NK cells.  
 CC !- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane  
 CC via its N-terminus.  
 CC !- SIMILARITY: Contains 1 cyclophilin-like PPIase domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC  
 CC EMBL; L04388; AAA35734.2; -.  
 CC EMBL; AF184110; AAD56402.1; -.  
 CC PIR; A47328; A47328.  
 CC HSSP; Q27450; 1A338.  
 CC Genew; HGNC:7833; NKTR.  
 CC MIM; 161565; -.  
 CC GO; GO:0046018; F:cyclosporin A binding; TAS.  
 CC InterPro; IPR002130; CSA\_PPIase.  
 CC Pfam; PF00160; pro\_isomerase; 1.  
 CC PRINTS; PR00153; CSAPPISMRASE.  
 CC PROSITE; PS00170; CSA\_PPIASE\_1; 1.  
 CC PROSITE; PS00072; CSA\_PPIASE\_2; 1.  
 CC Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.  
 CC DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.  
 CC FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).  
 CC FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).  
 CC FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).  
 CC FT DOMAIN 194 244 ARG/SER-RICH.  
 CC FT DOMAIN 466 574 ARG/SER-RICH.  
 CC FT DOMAIN 664 814 ARG/SER-RICH.  
 CC FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.  
 CC SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;  
 Query Match 3.2%; Score 166; DB 1; Length 1462;  
 Best Local Similarity 19.7%; Pred. No. 0.068;  
 Matches 136; Conservative 97; Mismatches 253; Indels 204; Gaps 28;  
 QY 39 VLDSKDAFVQFRVPPSSSGC--ALGRAGQHRGSATSLVFKQKTIITSMWDTKGIKTVES 96  
 DB 607 VIPLSDSPPSKWKPGQKPKPSYERIQEMAKTTHLLPIQSTY-SLANIK--ETGSSSS 663  
 QY 97 LHSKENNTREESMSSV-----QKDNFYQHNMEKLENV-SQLGFDKSPV 140  
 DB 664 YHKREKNSQSTVSKYSDRSSESPRSRGRSRYSRSTYRSRSLASHSRSPS 723  
 QY 141 EKG--TOYLKHQHTAAMCKW---QNEGPHSERLLESEPPAVTVLPQFNSNANTV 188  
 DB 724 SRSHRNKYSIHSCQSRSSSTSTSSDDGRKARLRSSGKNSVSHKHHSSSKTLHS 783  
 QY 189 -----DOSPKDDHSDTNS--EESRDNQQLTHVLANAKQWDEQGRARSHQKCGK 240  
 DB 784 KYVGRDRSSCVRKYSERSSLDYSSDSEQ--SSVQATQSAQEXE-KQGMERTHNQEK 840  
 QY 241 ACHPAEACAGCQOQETDVVSEPLSD--TGSEVDGTGLKNANRLNROESSIGNSPPPEKE 298  
 DB 841 -----NRGEKSKSEREPCHSKKTKLENLSDHLRNGSKPKRK-----NYAGSKWD 886







modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X55027; CAA38846.1; -;  
EMBL; U88551; AAC48720.1; -;  
EMBL; X55489; CAA39109.1; -;  
PIR; S15901; S15901.  
InterPro; IPR001819; Chromogranin\_AB.  
InterPro; IPR001990; Granin.  
Pfam; PF01271; Granin; 1.  
PRINTS; PR00659; CHROMOGRANIN.  
PROSITE; PS00422; GRANINS\_1; 1.  
PROSITE; PS00423; GRANINS\_2; 1.  
KW Sulfation; Cleavage on pair of basic residues; Signal.  
FT SIGNAL 1 20  
FT CHAIN 21 646 SECRETORANIN 1.  
FT PEPTIDE 418 484 GAWK PEPTIDE.  
FT PEPTIDE 634 646 SECRETOLYTIN.  
FT DISULFID 36 57 BY SIMILARITY.  
FT MOD RES 158 158 SULFATION (POTENTIAL).  
FT MOD RES 315 315 SULFATION (BY SIMILARITY).  
FT CONFLICT 64 64 N -> S (IN REF. 1).  
FT CONFLICT 70 70 N -> D (IN REF. 2).  
FT CONFLICT 93 98 SEAPGL -> PRSPRAS (IN REF. 3).  
FT CONFLICT 181 181 T -> M (IN REF. 2).  
FT CONFLICT 261 261 H -> R (IN REF. 2).  
FT CONFLICT 386 386 P -> R (IN REF. 2).  
FT CONFLICT 481 481 H -> L (IN REF. 3).  
FT CONFLICT 597 597 M -> V (IN REF. 2).  
SQ SEQUENCE 646 AA; 73339 MW; 420DB1178FD9E415 CRC64;  
  
Query Match 3.1%; Score 161.5; DB 1; Length 646;  
Best Local Similarity 20.7%; Pred. No. 0.043;  
Matches 158; Conservative 87; Mismatches 249; Indels 269; Gaps 39;  
  
QY 15 RWDAAATPPAASDARSPPGQRRLDS--KDAPVQFRVPPSSGCGALGRAGQHRGSATS 72  
DB 80 RFEVRLRLDPA--DTSEAPGLSSRE--DSGEGDAQVPTVADTESGHSRERAGEPPGSQVA 136  
QY 73 LVFQKQITSMWDYKINGIKTVESLSHSENNNTSESSNMSSVQKDNFYOHNMKLENVSQ 132  
DB 137 -----KAKYTKYSQEQNREEMVK-----YQ-KRERGEGVGESE 169  
QY 133 LGFDKSPVEKQYKIQHQTAAKQWQEGPHSERLLESEPPAVTLVPEQFSNANVDQSS 192  
DB 170 ERLSEGPGKAQTAFLNQ-----RNQTPAKKEELVSR-----YDTQSARGLEKSH 213  
QY 193 PKDHSNTNSBESRDNQOFLTHVLKLANAKQWMEDEQGEARSHQKCGKACHPAACAGCQ 252  
DB 214 SRERSQESGEBTKSQENW-----PQELQRHPEQGEA-----PGES----- 249  
QY 253 QEETDVVSESPLSDTSGSDVGTGLKNANRLNQESSLGNSSPPFEKESPESPMDVDNSKN 312  
DB 250 --EEDA--SEVD-----KHGRPRHHGRSRPDRSSQEGNPL----- 284  
QY 313 SCQDSEADETSPGDEQEDSSAQANKPGRFPQPEADTELKRSKAGGGEIRLHFQFE 372  
DB 285 -----EEESHVG-----TGSDEKARHPAFALEBGAEV-----GEEVRRHSA-- 324  
QY 373 GGESRAGNDVNAKRPSTSSINVECRNSKQGRKDKITDFHVRVPAEDKREKQCEMK 432  
DB 325 -----AQAPGDLQAGFGGRGRGEH-----QALRRPSEES--LEQENKR 361  
QY 433 HORTERKIPKYPHLSPPKQKWLGPPIEM--RRMPCRGIRLPLR-----PSANHTVIR 486  
DB 362 H-----GLSFDLNAQYSESEERGPAG---PSVRAGGEAAAYSTLGQ 405  
QY 487 VDLIR-IGEVKPPFTPHK---DLWDNKHVKMPCSEQNLVPVDENGERAAAGRWELIQT 542  
DB 406 TDEKFLGE-----THRVQESQDKARRLPGELRNLYDYGEEKGEAAAGKQW--- 455  
QY 543 ALLNRLTRPQ-NLKDAILLKYNVAISKWDFDALDFWKNVLEAEAAQHLQSYILPDWVKI 601

DB 456 -----POGDERDA-----DENREARLRG----- 474  
QY 602 ALCLPNTCTQPIPLKKNHSTMSQEQIASLILANAF-----CTFPRR----- 646  
DB 475 -----KOYAPHHT--EKRLGELL--NFFYDPSQWKSRRFERKDPMDSDSL 516  
QY 647 -----NAKMKSEYSSYPDINFN-----RLFE-----GRSSRKPEKIKTLFCYFRVATEKKP 692  
DB 517 EGBEENGLTINEKNFPPEYVNDWWEKPPFEEDVNWGYEKNPVKLDELKQYDRAVELDQ 576  
QY 693 TGLVTFTRQSLDEPPEWERCEKLLTRLUHVTYEGTIEGNGQML 735  
DB 577 --LIHYRKSAAE-FPDFYDSEEQMSPQH-TAENBEERKAGQGV 615  
  
RESULT 6  
DMPI\_HUMAN  
ID DMPI\_HUMAN STANDARD; PRT; 513 AA.  
AC Q13316; O43265;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix protein-1) (DMP-1).  
GN DMPI.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Molar;  
RX MEDLINE=97321043; PubMed=9177774;  
RA Hirst K.L., Simmons D., Feng J., Apelin H., Dixon M.J., McDougall M.;  
RT "Elucidation of the sequence and the genomic organization of the human dentin matrix acidic phosphoprotein 1 (DMPI) gene: exclusion of the locus from a causative role in the pathogenesis of dentinogenesis imperfecta type II.";  
RT Genomics 42:38-45(1997).  
RL [2]  
RN SEQUENCE FROM N.A. (ISOFORM 2).  
RP McDougall M., Juan X., Simmons D., Feng J.;  
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RX SEQUENCE OF 462-513 FROM N.A., AND GENE MAPPING.  
RX MEDLINE=96163890; PubMed=8586437;  
RA Apelin H.M., Hirst K.L., Crosby A.H., Dixon M.J.;  
RT "Mapping of the human dentin matrix acidic phosphoprotein gene (DMPI) to the dentinogenesis imperfecta type II critical region at chromosome 4q21.";  
RT Genomics 30:347-349(1995).  
RL [4]  
CC -!- FUNCTION: Could be involved in the induction of mineralization of extracellular matrix.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q13316-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q13316-2; Sequence=VSP\_004191;  
CC -!- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast, ameloblast and cementoblast.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; U65378; AAB87728.1; -.
DR MIM; 60980; -.
DR GO; GO:0005578; C:extracellular matrix; TAS.
DR GO; GO:0005509; F:calcium ion binding; TAS.
DR GO; GO:0005178; F:integrin binding; TAS.
DR GO; GO:0001503; P:ossification; TAS.
KW Biomineralization; Extracellular matrix; Signal; Alternative splicing.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 513 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.
FT CARBOHYD 25 25 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 426 426 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SITE 364 366 CELL ATTACHMENT SITE (POTENTIAL).
FT VARSPLIC 46 61 Missing (in isoform 2).
FT CONFLICT 69 69 /FTID:VSP 004191.
FT SEQUENCE 513 AA; 55782 MW; 2C1FE319A5D106F CRC64;

Query Match 3.1%; Score 159.5; DB 1; Length 513;
Best Local Similarity 20.6%; Pred. No. 0.042;
Matches 90; Conservative 70; Mismatches 155; Indels 121; Gaps 18;

QY 18 AAATSPPAASDARSFPGRRVLDSKAPVQFRVPSSSGALGRAGQHRGATSILVFQ 77
Db 146 ASEESAPOQDSAQDTTSERELNED-----RVDSPKEG-----GSTQ---RS 187

QY 78 KTTGW-----MDTKGIKTVSESLHSENNNTRBESMWSVQK-DNPFY 120
Db 188 ESEFHWVGSGDSESHGSDGSELDDEGMQSDPESIRSRGNRMNSAGMKSESGENSE 247

QY 121 QHNMEK-----LENVSLQGFQKSPVEKGTQYLKQHTAACKWQNEGPHSELSEPP 174
Db 248 QANTQDSGSGQLLEHPSPKIFPKRSISE-----EDDRSE--LDNNNT 287

QY 175 AVTLVPEQFSNAN---VDQSPKDDHSDTNSSESRDNOQFLTHVKLANAKQTMEDQGRE 231
Db 288 MEEVKSDSTENSNSRTGLSPRRDSKGDSDSEKEN-----LSQESQNV 334

QY 232 ARSHQKGVKACHPAECACQOE-----TIVVSEPSLDTGSEDVGTGLKN 278
Db 335 GPSSSESSQEAANLSSQNSSESQEEVVSERGNPNPTTSYVEDQEDSDSESDSHTLSH 394

QY 279 ANRLNQE---SSLGNSPPEKESEPESPMDVNSKN---SCDSEADEETSPGFDEQED 332
Db 395 SKSRESEQADSESSSLNFSSES-PESPEDENSSQEGLOSHSSAASQSESHSEDD 453

QY 333 SSSAQTANKPSRFPQPREADTELRKRSAGKGIRLHFQFEGGESRAGMNDVNAKRPGSTS 392
Db 454 SDSQDS---SRSEKDSNTE-SKSSSEEDGQLK-NIEIE-----SR 489

QY 393 SLNVECRNSQKGRKD 408
Db 490 KLTVDAYHNKPIGDQD 505

RESULT 7
UN89_CABEL
ID UN89_CABEL STANDARD; PRT; 6632 AA.
AC AC 001761; Q17362;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.

```

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN-Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
transduction domains.";
RT J. Cell Biol. 132:835-848 (1996).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RT Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U33058; AAB00542.1; -.
DR EMBL; AF003131; AAB54132.2; -.
DR PDB; 1PHO; 20-DEC-00.
DR WormPep; C09D1.1; CE30426
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGC2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS50835; IG-LIKE; 49.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.

```

```

FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBROECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).

```

```

FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEY -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;

Query Match 3.1%; Score 159; DB 1; Length 6632;
Beat Local Similarity 20.2%; Pred. No. 1.2;
Matches 185; Conservative 136; Mismatches 387; Indels 206; Gaps 45;

Qy 9 PCTKRPWDAAATSPPAASDARSFPGQRRVLOSADAPQVFRVPPSSGGCALGRAGOHGR 68
Db 1407 PKKSPASPTKCKKSPAAEBVKSPKKEK---SPSPPTKCKKSPSPPTKKTGDEVKEKS 1462
Qy 69 SATSLVEKQKTIISWMDTK-----GIKTVESSE-SLHSEKNNNTRE-----ESM 110
Db 1463 PPKSPPTKCKKSPKPEPDVKSPPVKKEKSPDNTNIVESSSETIETKTETMTMTMTHESSES 1522
Qy 111 MSSVQKDNFYQHNMKLEKLVNSQLGFDKSPVEKGTQYLYKHQHTAAMCKWQNEGPHSERILLE 170
Db 1523 RTSVKKEKTPKVDKPKSPKPK--DKSPKSIITEIKSPVK-----KEKSPKVEKPK 1574
Qy 171 SEPPAVTLVPEQSFNANVTQSSPKDDHSDNBSERDNOQOFLTHVLANAKQWME-DEQG 229
Db 1575 ASPTKKESPEK--PASPTKSENEVKSPTKCKKSPK--KSVVEELKSPKESPKKADDPK 1631
Qy 230 REARSHQKCGKACHPAECACAGCQOEET-DVVSFSLSDTGSSEVGT-----GLKANRLN 283
Db 1632 KSPTKKESPEKSATEDVKSPTKCKKSPKVEKPTSPTKKESPTKKTDDDEVKSPKKE 1691
Qy 284 QKSSSLGNSP--PPEKSESESPMDVNSKNSQDSADBEFSPGDEQDESSSAQTANK 341
Db 1692 KSPQTVEEKPASPTKCKKSPK-----KSVVEEVKSPKESPEKABEKPSPTKKEKS 1743
Qy 342 PSRQPREADTELK-RSSAKGGHRLHFOFEGESRAGNDVNAKRP-----GSTSLNV 396
Db 1744 PEKSAABEVKSPTKCKKSPKSAEBKSPKCKESSPVKMADEVKSPKCKESPEKVEE 1803
Qy 397 ECRNSKHQHKDSKITDHFMRVP-----KAEDKKKEOC-EMKHQORTERKIPKY 443
Db 1804 KPASPTKCKKTPKESAAEBELKSPKCKKESSPSPKTKGDSKESKSPKESKSPKSPKPK 1863
Qy 444 IPPHLSPKKWLGTPIBEMRMPCGIRLPLRPSANHTVIRVDLLRIGEVKPKPPTHF 503
Db 1864 SPPG-SPKCKKSPK--EAEKPP-----APKLRDLKLQTVNKTDLAHL-EVVVEHATEC 1914
Qy 504 KDLWDNKHVMPCEQNLVPEDENGE-----RAKSRWELIQT 543
Db 1915 KWFLDGKEI---TTAQGVTVSKDDQFECRCSIDTTFMGSGTVSVVANAAGS-----VETK 1967
Qy 544 L-LNRLTRPQNLKDAILKYNVAYSKWDFL-ALIDF-----WDKVLSEAEQAH--LYO--- 592
Db 1968 TELKVLTPKE-----TKKEFTDKLEDMEVTKGDTVMQDVIALHSLYKWKYQ 2015
Qy 593 -----SILPDMVKIALCLPNICTQPIPLLKQKNHSGITMSQEQIASLANAFF 640
Db 2016 NGNLLEDGKGVTKNEENKSLIIPN--AQDSKITVEASNEVG-SSESSAQLTVPNPPS 2072
Qy 641 CT-----FPR-----RNAKMKSEYSSYP-----DINFNRLPFGSRSSKPEKLTFLC 682
Db 2073 TTPIVWDGPKSVTIKETETAEFAKATISGFPAPVKTWIN-EKIVE--ESITITTIKTEDV 2129
Qy 683 YFRFRTBKK--PTGLVTFTRQSL--EDFPWE-----RCEKLLTRL--HVTYEG---- 725
Db 2130 YTLKISNAKIEQGTGVKVTQAQNSAGQDSKQADLKVEPNVNAKPKSKQLTDKVADEGFLR 2189

```

```

QY 726 ---TTIEGNGQGLQVDFANRFVGGVTSAGLVQVEIRFLINPELIVSRFLTEVLDRNE-C 781
Db 2190 WNLELDGSPG-----TEVSWLNGQPLTKSTVQVWDHGDGT 2227
QY 782 LIITGTEQYSEYTG 795
Db 2228 YHVITAEAKPEMSG 2241

RESULT 8
MDN1_HUMAN
ID MDN1_HUMAN STANDARD; PRT; 5596 AA.
AC Q9NU22; O15019;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Midasin (MIDAS-containing protein).
GN MDN1 OR KIAA0301.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC PubMed=12102729;
RA Garbarino J.E., Gibbons I.R.;
RT "Expression and genomic analysis of midasin, a novel and highly
RT conserved AAA protein distantly related to dynein.";
RL BMC Genomics 3:18-18(2002).
RN [2]
RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.
RC TISSUE=Testis;
RA Tracey A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE OF 3550-5596 FROM N.A.
RC TISSUE=Brain;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kohani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC the assembly/dissassembly of macromolecular complexes in the
CC nucleus.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Contains 1 WWFA domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF503925; AAM7722.1; -
CC EMBL; AL096678; CAB86660.1; -
CC EMBL; AL096678; CAB86661.1; -
CC EMBL; AB002299; BAA20761.1; -
CC Genew; HGNC:18302; MDN1.
CC GO; GO:0005634; C:nucleus; NAS
CC GO; GO:0016887; F:ATPase activity; NAS.
CC GO; GO:0003754; F:chaperone activity; NAS.
CC GO; GO:0006461; P:protein complex assembly; NAS.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC Pfam; PF00004; AAA; 1.
CC SMART; SM00382; AAA; 7.

```

```

DR SMART; SM00327; VWA; 1.
DR PROSITE; PS0234; VWFA; 1.
KW Chaperone; ATP-binding; Repeat; Nuclear protein.
FT NP_BIND 329 336 ATP (POTENTIAL).
FT NP_BIND 677 684 ATP (POTENTIAL).
FT NP_BIND 1084 1091 ATP (POTENTIAL).
FT NP_BIND 1390 1397 ATP (POTENTIAL).
FT NP_BIND 1753 1760 ATP (POTENTIAL).
FT NP_BIND 2066 2073 ATP (POTENTIAL).
FT DOMAIN 3566 3573 POLY-GLU.
FT DOMAIN 4784 4791 POLY-GLU.
FT DOMAIN 5008 5013 POLY-GLU.
FT DOMAIN 5182 5187 POLY-GLU.
FT DOMAIN 5384 5583 VWFA.
FT CONFLICT 2287 2312 RFLSDMPVHGDISRAMNRGLIYI -> S (IN REF.
SQ SEQUENCE 5596 AA; 632802 MW; 586C62616A1F96D4 CRC64;

Query Match 3.1%; Score 158.5; DB 1; Length 5596;
Best Local Similarity 18.5%; Pred. No. 1;
Matches 174; Conservative 142; Mismatches 312; Indels 313; Gaps 45;

QY 90 KIVSESLHSKENNTREESMVQKONFYQHNNKLENVSLQGFDKSPVEKGYLQK 149
Db 4872 KYPEPEALDLPDLNLSDEKNGGSDTDN-----EGEENPLFIKKEPEAGHEAER 4925
QY 150 HQTAMCKWQ-----EGPHERLSEPPAVTLVPRQFSNANVDQSSPKDDHSDT 200
Db 4926 GETETD---QNESQSPQPEEGPSDDKAGE-----EMDTGADDQDDAAQHPEE 4974
QY 201 NSESRDNQQLTHVYKANAK-----QTMEDQGREARSHQKQKACHPAEACA 249
Db 4975 HSEEQ--QSVEEKDKADEGGGNGPADQGFQPEEEREDSDTEQVPEALERKEH-A 5031
QY 250 GCOQ-----BETDVVSEPLSDTGSVDVGTGLKNANRLNRQESSI----- 289
Db 5032 SCGQTGVENMQTQAMELAGAAPEKEGKEHSGAADAQAEGHESFIAQLASOKHTR 5091
QY 290 GNSPPFEKESPESEPMVDVNSKNSCQDSEADEETSPGDEQEDSSSAQTANKPSFQPRE 349
Db 5092 KNTQSFKKK-----PGQADNERSMGDNHVRVHKRLTVD--TDSHAEGQPAQPAQVED 5144
QY 350 ADT--ELKRSAGGEIRLHFQFEGGSRAGMNDVNAKRPGSTSLNVCEKNSQKGRK 407
Db 5145 ADAFEHIKQGSDAYD-----AQTYDVASK-----EQQSAK 5175
QY 408 DSKITDHPMVPKAEKKEQCEMKHQRTKIPKYPHLSPPDKKWLGTPIEEMRMRP 467
Db 5176 DS-----GKQDEEBIEDTLMDEHQ-----EEFKAADVEQLK---- 5208
QY 468 CGIRLPPLRPSANHTVTIRVDLRLRIGVEPKFPPTHFKDLWNKHVKMPCSEONLYEVEDE 527
Db 5209 -----PEEIKSGTAPLGFDEMEV-ELQTVKTEEDQD-----PRDKAKHETENE 5252
QY 528 NGERAAGSRWELIOTA---LNLRLTRPNLKDAILKYNVAYKKWDFALIDFWD----- 579
Db 5253 KPER---SRESTIHTAHQFLMDTIFQP-FLKD-----VNELRQELERQLEMMQWPRESG 5301
QY 580 KVLSEEAQHLVYOSTLPDMVKIALCLPNICQPIPLLKQKNHNSITMSQEQIASLANAF 639
Db 5302 NPEBEKVAEAMWQSYL-----ILTAP--LSQRICEELRLILE----- 5336
QY 640 FCTFPRRNAMKSEYSSYPDINFNRLPEGRSSR-----KPEKLTLCFYFRVT 688
Db 5337 ----PTQAAKLKGDYRTCKRLNIRKIPYIASQFKDKIWLRTKPSKRYQICL---AI 5389
QY 689 EKKPTGLVTFTRQSLDFPPEWERCEKLLRLHVTYEG-TIEGNGQGLQVDFANRFVGG 747
Db 5390 DSSSMVDNHTKQ-----LAFESLAVIGNALTILEV-----G 5421
QY 748 VTSAGLVQVEIRFLINPELIVSRFLTEVLDRNECLITGTQYSEYTCYATYWARSH 807
Db 5422 QIAVCSFGSVK-LLHP-----FHEQFSDYSG-SQLRLCK---- 5455

```

QY	808	DRSERDDQWRRTEI	VAIDALHFRFYLDQFVPEKIRRELKAYCGFLRPGVSENLSAVA	867
		: : : :	: : : :	
Db	5456	-----FQKKTKIA	-----QFL-BSVANMF-AAAQQLSNISSETAQLLL	5493
		: : : :	: : : :	
QY	868	TGNWCGCAGFGDARL	KALIOILAAVAERDV--VVFTEGSELMRDIYSMHFLTERKL	924
		: : : :	: : : :	
Db	5494	WSDGRLGF-----	LEGERVLAQVAQRANITFVFLVDNPSRD-----SILDIKV	5542
		: : : :	: : : :	
QY	925	TV-----GEVYKLILRYNEECRNGSTPGDIKLYPFYIYHAV	961	
		: : : :	: : : :	
Db	5543	PIPKGPGEMPE--IRSYMEZ-----	FFPYIYII	5568
		: : : :	: : : :	

```

RESULT 9
ELAI_MOUSE
ID      ELAI_MOUSE      STANDARD;      PRT;      773 AA.
AC      QRCB77; Q80VB2; Q9R0Q5;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Transcription elongation factor B polypeptide 3 (RNA polymerase II
DE      DE transcription factor SIII subunit A) (SIII p110) (Elongin A) (EloA)
DE      DE Transcription factor SIII subunit A) (SIII p110) (Elongin A) (EloA)
DE      DE (Elongin 110 kDa subunit).
GN      TCEB3.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20044622; PubMed=10575222;
RT      Aso T., Aminoto K., Takebayashi S., Okumura K., Hatakeyama M.;
RT      "Structural organization and chromosome location of the mouse elongin
RT      A gene (Tceb3)";
RT      A gene (Tceb3)";
PL      Cytogenet. Cell Genet. 86:259-262(1999).

```

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Bone;  
RX MEDLINE=22354683; PubMed=12466851;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Knapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chochia J., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmerond S., Gustintchik S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKensie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Maeyashiki A., Yoshino M., Waterston R., Lander E.S., Rogers J.,

RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";   
RL Nature 420:563-573 (2002).

RN	[3]	SEQUENCE FROM N.A.
RP		STRAIN=FVB/N; TISSUE=Breast tumor;
RC		RX
		MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Iqouellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smallos D.E.,  
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16895-16903 (2002).  
 CC -I- FUNCTION: Sill, also known as elongin, is a general transcription  
 CC elongation factor that increases the RNA polymerase II  
 CC transcription elongation past template-encoded arresting sites.  
 CC Subunit A is transcriptionally active and its transcription  
 CC activity is strongly enhanced by binding to the dimeric complex of  
 CC the SIII regulatory subunits B and C (elongin BC complex) (By  
 CC similarity).  
 CC -I- SUBUNIT: Heterotrimer of an A (A1, A2 or A3), B and C subunit (By  
 CC similarity).  
 CC -I- DOMAIN: The elongin BC complex binding domain is also known as BC-  
 CC box with the consensus [APST]-L-x(3)-C-x(3)-[AIVL].  
 CC -I- SIMILARITY: Contains 1 F-box domain.  
 CC -I- SIMILARITY: Contains 1 TFS2-N domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

[illegible]

	Query Match	3.0%	Score 156.5	DB 1	Length 773
Best Local Similarity	20.6%	Pred. No. 0.11			
Matches 124	Conservative	78	Mismatches	204	Indels 195; Gaps 31
Qy	79	TITSMWDTGKIKTVESLSHSENNNTRE	-----ESMSSSVQKDNFYQHN	123	
		:::	:::	:::	
Qy	39	TVDTILVETGVGKTVNSFRGHEQVGNFARDLVAQWKGLVPVERNSEADQDEKNSRRP	98		
		:::	:::	:::	

```
QY 124 MEKLENSVQL--GFDKSPVEKGTQ-YLKQHTAAMCKM-QNEGPHS-----ERL 168
DB 99 RDALQREELQYQESWKPSSGSRYSPEHQKCKKLSEPERPHKVAHSEKDERKAC 158
QY 169 LSEBPATVLPVQESNANVQ---SSPKDDSDTNSERSRNOQFLTHVKLANA-KQTM 224
DB 159 HKVSP-PYSSDPESDYGHVQPPSPSSPHQMTDLSRSPEDQEPHIIHQKPGKVSHTF 217
QY 225 ED-----EGRERARSHQK---CGKACHPAEACAGCQCBETDVSPLSDTGSDEV 272
DB 218 QDRLGVLHGEQKGAHVSHKHQSRSHKHPADA---REDEKISAVSRE----- 264
QY 273 GTGLKXNANRLNRQESLGNSSPPFEKESPEPMDVNSKNSQDSRADETSFGFDEQED 332
DB 265 ---KSHKASSKEESRLLSGDSAKEKLPSSVVKDKDR---EGSLKKKFSPLDVAAD 317
QY 333 SSSAQTANKPSFPQREADETELKRS-----SAKG-----GHIRLHFQEGGESR- 377
DB 318 NH-----FKPKPKHDKSEKAKSKNKSQVGDVSGRTGDTPLPKAKEKVPNHLKAQEGKVRT 373
QY 378 ---AGM-----NDVNV-----AKRPGSTSLNVECRNSK 402
DB 374 NADGKSAGLLHPKAEETDDEFEERTMSFESVLSYDQPRKKKKVVKTSSTALGEKGLK 433
QY 403 QHGRKDSKIT---DHFMRVPKAEKDKQCEM---KHQTERKIP----- 441
DB 434 ---KKDSKSTKNLNSAQKLPKNWENKSEKLPAGAEPTPRKVPDTPLPALPDIPALP 490
QY 442 ---KYIP-----PHLSPDKKWLGTPIEM-----RRM-----PROGIRLPPLP 477
DB 491 HANYRPLPSLELIPSPQPKKAFSSPQEEBAGFTGRRMNSKMQVYSGSKCAY-LPRM-- 547
QY 478 SANHTVTIRV-----DLLRIGEVF-----KPPFTFKLWMDNKHVMQPCSEONLYPVEDE 527
DB 548 MTLHQQCIRLVKNNDISIFEVGVFVSIVLEP-----VLERTCPDQLYRIESC 594
QY 528 N 528
DB 595 N 595

RESULT 10
MINT MOUSE STANDARD; PRT; 3644 AA.
AC Q62504; Q80TN9; Q99PS4; Q9Q2W2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
[1]
SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
TISSUE=Testis;
RX MEDLINE=99379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel msx2 binding protein, recognizes and
regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10578-10590(1999).
[2]
SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Horii F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Kono H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
RA Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watanishi A., Muramatsu M., Hayashizaki Y.;
```

```
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
J/Su(H) activity.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 318-578 FROM N.A.
RN
RC TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet F., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fizesac A., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
unconventional type I myosins.";
RL Genomics 40:332-341(1997).
[5]
RP SEQUENCE OF 2598-3644 FROM N.A.
RN
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
II. The complete nucleotide sequences of 400 mouse KIAA-homologous
cDNAs identified by screening of terminal sequences of cDNA clones
randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
[6]
RP TISSUE SPECIFICITY.
RX MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostezka U., Astrahantseff K., Bourteelle S., Dillinger K.,
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
pathway.";
RL EMBO J. 21:5417-5426(2002).
[7]
RP FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=22483652; PubMed=12594956;
RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
of Notch/RBP-J signalling pathway.";
RL Immunity 18:301-312(2003).
CC -!- FUNCTION: Essential corepressor protein, which probably regulates
different key pathways such as the Notch pathway. Negative
regulator of the Notch pathway via its interaction with RBPJ, and
which prevents the association between Notch1 and RBPJ, and
therefore suppresses the transactivation activity of Notch
signaling. Blocks the differentiation of precursor B cells into
marginal zone B cells. Probably represses transcription via the
recruitment of large complexes containing histone deacetylase
proteins. May bind both to DNA and RNA.
CC -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and
MTAIL1. Interacts with the nuclear receptors RAR and RXR.
Interacts with RAR in absence of ligand. Bind to the steroid
receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
Interacts with RBPJ; this interaction may prevent the
interaction between RBPJ and NOTCH1.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62504-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62504-2; Sequence=VSP_008564;
CC Note=No experimental confirmation available;
CC Note=TISSUE SPECIFICITY: Highly expressed in testis. Expressed at lower
level in brain, lung, spleen, liver and kidney. Weakly expressed
in cardiac and skeletal muscles and ovary. In spleen, it is
```

expressed in follicular B-cells, while it is weakly expressed in marginal zone B-cells.

-1- DOMAIN: The RID domain mediates the interaction with nuclear receptors.

-1- DOMAIN: The SPOC domain, which mediates the interaction with NCOR2, is essential for the repressive activity (By similarity).

-1- SIMILARITY: Belongs to the Spen family.

-1- SIMILARITY: Contains 1 RIR (receptor interacting) domain.

-1- SIMILARITY: Contains 4 RIR (receptor interacting) domains.

-1- SIMILARITY: Contains 1 SPOC domain.

-1- CAUTION: Ref.4 sequence differs from that shown due to multiple frameshifts and conflicts that create stop codons.

-1- CAUTION: Ref.5 sequence differs from that shown due to what seems to be the presence of intronic sequence in the cDNA.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

-----  
EMBL; BY26481; ; NOT\_ANNOTATED CDS.

DR EMBL; AF156529; AAD55931.1; ALT\_INIT.

DR EMBL; AB055980; BAB32786.1; -

DR EMBL; Z78160; CAB01562.1; ALT\_SEQ.

DR EMBL; AK122402; BAC65684.2; ALT\_SEQ.

DR MGD; MGI:1891706; Mint.

DR InterPro; IPR000504; RNA\_rec\_mot.

DR Pfam; PF000076; Rrm; 3.

DR SMART; SM00360; RRM; 3.

DR PROSITE; PS01012; RRM; 4.

DR PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.

DR PROSITE; PS00910; SPOC; 1.

-----  
KW Transcription regulation; Repressor; Nuclear protein; DNA-binding; DNA-binding; Coiled coil; Coiled coil; Alternative splicing; Polymorphism.

-----  
FT DOMAIN 1 574

FT RNA-BINDING (RRM) 1.

FT RNA-BINDING (RRM) 2.

FT RNA-BINDING (RRM) 3.

FT RNA-BINDING (RRM) 4.

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT COILED COIL (POTENTIAL).

FT CONFLICT 1815 1815 A -> V (IN REF. 3).  
FT CONFLICT 2097 2097 G -> A (IN REF. 3).  
FT CONFLICT 2201 2202 MISSING (IN REF. 3).  
FT CONFLICT 2322 2322 A -> V (IN REF. 3).  
FT CONFLICT 2385 2385 P -> Q (IN REF. 3).  
FT CONFLICT 2502 2502 R -> K (IN REF. 3).  
FT CONFLICT 2505 2505 E -> K (IN REF. 3).  
FT CONFLICT 2519 2519 D -> N (IN REF. 3).  
FT CONFLICT 2554 2554 T -> S (IN REF. 3).  
FT CONFLICT 2679 2688 LVSTPAGPVN -> VGEHPWARD (IN REF. 3).  
FT CONFLICT 3010 3010 L -> P (IN REF. 3 AND 5).  
SQ SEQUENCE 3644 AA; 398750 MW; 9C7EC49A81A7DA4A CRC64;

Query Match 3.0%; Score 156; DB 1; Length 3644;

Best Local Similarity 19.2%; Pred. No. 0.83;

Matches 128; Conservative 86; Mismatches 203; Indels 250; Gaps 31;

QY 15 RWDAAATSPAAADARSFPG-----RQRRVLDKSDAPVQPRVPPSSSGCALGRAGQH 66

Db 1354 RWDPMKQDSRFV-SFNSVTKRSLKRSVRDLEPGV-----PSDS-----DEDAEH 1403

QY 67 RGSATSLVFQKQTTITSMMDTKGKTVESLSKKN-NTRTESMSSSVQKDNFYQHME 125

Db 1404 RS-----QSPRASFYDSPRL-----SFLLRDQKLRERDERLASSLERNFYSALD 1452

QY 126 KL-----ENVQLGF-----DKSPVEKGTQ-----YLKQ 149

Db 1453 KTTTPDTKALLERAKSLSSSSREENW3FLDMSDFANFRNNKDKKVDASAPRIPSWYMK 1512

QY 150 HQTAAMCKWNEGPHSERLESEPPAVTLVPEQFSNANV-DQSPKDDHSDTINSEES--- 205

Db 1513 KKTIDSEGLDKKDER-REEQEQELFASPLHSSIFEQDSKRLQHLERKSESDFP 1571

QY 206 -----RDNQO-----FLTHVKL-----ANAKOTMEDEQGRARSHQCKGA 241

Db 1572 PGRLYGRQSEGANSTSDSVQEPVWLFHRSFMBILTRMQQKEKDKQKAEKQEE--PE 1629

QY 242 CHP-----AEACAGCQ-----

Db 1630 THPTPEPAATKEPEKAPVSAGLPAVTITVTTPPEPASSAPEKAEAAAPSPAGEKPA 1689

QY 253 -----QETDWSSES-----PLSDTGSSEVGTGLNANRLNRQESSLGNSSPPFEK----- 297

Db 1690 EPAPVSETKLVSEPVSVFVEQPRQSDVPPG--EDSEDSQDSAAALAPSAQESAATDAPV 1747

QY 298 -----ESPESEPMVDNKNKSCDSEADE-----ETSPGFD--- 328

Db 1748 CVNAEPLTPGTTVSQVSESSVDKPPSPQL--SKLTORSEAEKGVKPKPTTPTSTPDA 1805

QY 329 -----EQED-----SSSAQTANKPSRFQPREA 350

Db 1806 TQNAVASEAQPPASEDVANPPVAADKRTNKSKRTSVQAAAASVVEKPVTRKSERI 1865

QY 351 DTELKRSKAGGEIR--LHFQEGGE-----SPAGMNDVNAKPGSTSSINVECRNSKQ 403

Db 1866 DREKLKSSSPRGEAQKLLKWEAEKTRTASKSGSDTIEHPPEPLSL-----RGRRR 1920

QY 404 HGRK-DSKITDHFMRVPAEDKKEQCEMKHQRTKIPKIPKIPKIPKIPKIPKIPKIPKIP 462

Db 1921 NVRSVYATMTDHSRSPAKEP--VEQPRVTRKRLERLEQEAUVPTTTPRR---GRPKTR 1975

QY 463 RMPRCG 469

Db 1976 RRAEDG 1982

RESULT 11

MINT\_HUMAN

ID MINT\_HUMAN STANDARD; PRT; 3664 AA.

AC Q96T58; Q9H9A8; Q9NMH5; Q9UQ01; Q9Y556;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).  
GN MINT OR SHARP OR KIA0929.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI TaxID=9606;  
RN [1] \_  
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,  
RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBP4; MBD3;  
RP RAR AND MTA1L1.  
RC TISSUE=Liver, and Pituitary;  
RX MEDLINE=21231190; PubMed=11331609;  
RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,  
RA Hon M., Evans R.M.;  
RT "Sharp, an inducible cofactor that integrates nuclear receptor  
RT repression and activation.";  
RL Genes Dev. 15:1140-1151(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Bird C.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 294-3664 FROM N.A.  
RA Rhodes S., Huckle E.;  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.  
RC TISSUE=Embryo, and Teratocarcinoma;  
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
RA Arita M., Nakamura Y., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 2002-3664 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99246063; PubMed=10231032;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIII.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:63-70(1999).  
RN [6]  
RP INTERACTION WITH PPARD.  
RX MEDLINE=21874127; PubMed=11867749;  
RA Shi Y., Hon M., Evans R.M.;  
RT "The proximal proliferator-activated receptor delta, an integrator  
RT of transcriptional repression and nuclear receptor signaling.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).  
RN [7]  
RP FUNCTION, AND INTERACTION WITH RBPSUH.  
RX MEDLINE=2261914; PubMed=12374742;  
RA Oswald F., Kosterka U., Aszrahantseff K., Bourteelle S., Dillinger K.,  
RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptray S.,  
RA Schmid R.M.;  
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling  
RT pathway.";  
RL EMBO J. 21:5417-5426(2002).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.  
RX MEDLINE=22777836; PubMed=12897056;  
RA Ariyoshi M., Schwabe J.W.R.;  
RT "A conserved structural motif reveals the essential transcriptional  
RT repression function of Spen proteins and their role in developmental  
RT signaling.";  
RL Genes Dev. 17:1909-1920(2003).  
CC -!- FUNCTION: Essential corepressor protein, which probably regulates  
CC different key pathways such as the Notch pathway. Negative  
CC regulator of the Notch pathway via its interaction with RBPSUH,  
CC which prevents the association between NOTCH1 and RBPSUH, and

CC therefore suppresses the transactivation activity of Notch  
CC signaling. Blocks the differentiation of precursor B cells into  
CC marginal zone B cells. Probably represses transcription via the  
CC recruitment of large complexes containing histone deacetylase  
CC proteins. May bind both to DNA and RNA.  
CC -!- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with  
CC NCOR2, HDAC1, HDAC2, RBP4, MBD3 and MTA1L1. Interacts with  
CC RBPSUH; this interaction may prevent the interaction between  
CC RBPSUH and NOTCH1. Interacts with the nuclear receptors RAR and  
CC PPARD. Interacts with RAR in absence of ligand. Bind to the  
CC steroid receptor RNA coactivator SRA.  
CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.  
CC -!- TISSUE SPECIFICITY: Expressed at high level in brain, testis,  
CC spleen and thymus. Expressed at intermediate level in kidney,  
CC liver, mammary gland and skin.  
CC -!- INDUCTION: By hormone 17-beta-estradiol (E2).  
CC -!- DOMAIN: The RID domain mediates the interaction with nuclear  
CC receptors (By similarity).  
CC -!- DOMAIN: The SPOC domain, which mediates the interaction with  
CC NCOR2, is essential for the repressive activity.  
CC -!- SIMILARITY: Belongs to the Spen family.  
CC -!- SIMILARITY: Contains 1 RID (receptor interacting) domain.  
CC -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.  
CC -!- SIMILARITY: Contains 1 SPOC domain.  
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous  
CC gene model prediction.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AF356524; AAK52750.1; --  
CC EMBL; AL034555; CAB85442.1; ALT SEQ.  
CC EMBL; AL034555; CAB85444.1; ALT SEQ.  
CC EMBL; AL450998; -- NOT ANNOTATED CDS.  
CC EMBL; AL036858; CAB51072.1; ALT INIT.  
CC EMBL; AK000882; BAA91405.1; ALT INIT.  
CC EMBL; AK022949; BAB14324.1; ALT INIT.  
CC EMBL; AB023146; BAA76773.1; --  
CC InterPro; IPR000504; RNA\_rec\_mot.  
CC PDB; 1OM1; 19-AUG-03.  
CC Pfam; PF00076; rrm; 4.  
CC SMART; SM00360; RRM; 4.  
CC PROSITE; PS0102; RRM; 4.  
CC PROSITE; PS0030; RRM\_RNP\_1; FALSE\_NEG.  
CC PROSITE; PS0917; SPOC; 1.  
CC Transcription regulation; Repressor; Nuclear protein; DNA-binding;  
CC Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.  
CC KW Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.  
CC FT DOMAIN 1 573  
CC FT DOMAIN 6 81  
CC FT DOMAIN 335 415  
CC FT DOMAIN 438 513  
CC FT DOMAIN 517 589  
CC FT DOMAIN 688 715  
CC FT DOMAIN 977 1004  
CC FT DOMAIN 1170 1191  
CC FT DOMAIN 1408 1428  
CC FT DOMAIN 1496 1529  
CC FT DOMAIN 1592 1612  
CC FT DOMAIN 1928 1944  
CC FT DOMAIN 2201 2207  
CC FT DOMAIN 3498 3664  
CC FT DOMAIN 2130 2464  
CC FT DOMAIN 2709 2870  
CC FT DOMAIN 125 277  
CC FT DOMAIN 240 325  
CC FT DOMAIN 616 810  
CC FT DOMAIN 624 697  
CC FT DOMAIN 2428 2520  
CC INTERACTION WITH MSX2 (BY SIMILARITY).  
CC INTERACTION WITH RBPSUH (BY SIMILARITY).  
CC ARG-RICH.  
CC ARG-RICH.  
CC TYR-RICH.  
CC PRO-RICH.



FT DOMAIN 3220 3482 PRO-RICH.  
 FT VARIANT 970 970 A -> V (in dbSNP:848208).  
 FT 1091 1091 /FTID=VAR\_017119.  
 FT 1091 1091 L -> P (in dbSNP:848209).  
 FT 2360 2360 /FTID=VAR\_017120.  
 FT 2360 2360 N -> D (in dbSNP:848210).  
 FT 956 956 /FTID=VAR\_017121.  
 FT CONFLICT 956 956 G -> D (in REF. 4).  
 FT SEQUENCE 3664 AA; 402245 MW; 5228C58533E5B27B CRC64;  
 Query Match 3.0%; Score 156; DB 1; Length 3664;  
 Best Local Similarity 18.8%; Pred. No. 0.84;  
 Matches 101; Conservative 78; Mismatches 232; Indels 126; Gaps 19;  
 QY 4 GPCEPCTKPRWDAATSPPAADARSFPGRQRRVLDKDAVQVRVPSPSSGCGALGRA 63  
 Db 1994 GPQCKGKNEPKVD--ATREATTVEV---GPQIGVKESMEEP---KAAEEAEGSQKRD 2044  
 QY 64 GQHRGSATSLVFKQKITISMDTKGKTVESELSHSEKNNNTREESMMSSVQKDNFYQHN 123  
 Db 2045 RKDAG-----TDKNPPETAPVEVWEKXPAPEKNKSKRGRSRNSGLAYD-----K 2089  
 QY 124 MEKLVNSQLGFKSPVEKGTQYLKQHOTAAKCKWQNEGPHSE-----RLIESEPPAVTLV 179  
 Db 2090 SASLKNVDAVSPRGAQAQAGE--RESGVAVPEKSESQKEDGLSSQLKSDPVPDK 2147  
 QY 180 PEQFSNANVDQSPKDDHSDTNSSESDNQOFLTHV--KLANAKQTMEDQEGREARSHQKC 238  
 Db 2148 PEK---EDVASGSPSE--ATLAKQMELEQAVEHIAKLAES-----2185  
 QY 239 GKACHAPACAGCQOQETDVVSPLSDTSGEDVGTGLKN--ANRLNQESSLGNSPPFEK 297  
 Db 2186 ASAYKADAPGLAPEDRD----KPAHQASETEALAAIGSIINDISGEPPNFAPPPYPG 2241  
 QY 298 ESE-----PESPMVDNSKNSQDSEADETSPGFDEQEDSSSAQTANKPS 343  
 Db 2242 ESQTDLOPPAGAQALQPSSEGMETDEAVSGILETEATESSRPVNPAPDPSAGPTDK 2301  
 QY 344 RFOPREADTELKRSKAKGGEIRLHFQEG-----373  
 Db 2302 RGNSSSETSHVPEAKSGKEVEVTLVRKDKGRQKTRSRKRNTNKKVAPVESHVPESNQ 2361  
 QY 374 --GESRAGMDVNAKRPGSTSSNLNVECRNSQKGRKDKSKITDHFMVVPKADKKEQCEM 431  
 Db 2362 AQGESPAANGTTVQHPAPQEQ---EKQSEKPHSTPPQSTSLDKIPSTENSSQELISVE 2418  
 QY 432 KQORTERKIPKYPHLSPPKMKLGTPIEMRRMPCRGIRL-----PPLRPS 478  
 Db 2419 ERPTTKASVPDLPPPPQ-----APVDE---EPQARFRVHSTIESDPVTPPSDPS 2466

RESULT 12  
 ID DMP1 RAT STANDARD; PRT; 489 AA.  
 AC P38193;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Dentin matrix acidic phosphoprotein 1 precursor (Dentin matrix  
 DE protein-1) (DMP-1) (AG1).  
 GN DMP1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Tooth;  
 RX MEDLINE=93286101; PubMed=8509401;  
 RA George A., Sabay B., Simonian P.A., Veis A.;  
 RT "Characterization of a novel dentin matrix acidic phosphoprotein.  
 RT Implications for induction of biomineralization.";  
 RL J. Biol. Chem. 268:12624-12630(1993).

CC -!- FUNCTION: Could be involved in the induction of mineralization of  
 CC extracellular matrix.  
 CC -!- TISSUE SPECIFICITY: Expressed in tooth particularly in odontoblast  
 CC and ameloblast.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L11354; -; NOT ANNOTATED CDS.  
 KW Biomineralization; Extracellular matrix; Signal.  
 FT SIGNAL 1 16 POTENTIAL.  
 FT CHAIN 17 489 DENTIN MATRIX ACIDIC PHOSPHOPROTEIN 1.  
 FT DOMAIN 41 44 POLY-PRO.  
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT SITE 334 336 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SEQUENCE 489 AA; 53058 MW; 59F8381479DDA085 CRC64;  
 Query Match 3.0%; Score 155; DB 1; Length 489;  
 Best Local Similarity 21.1%; Pred. No. 0.071;  
 Matches 78; Conservative 71; Mismatches 136; Indels 84; Gaps 17;  
 QY 94 SLSLHSEKNNNTREESMMS--SVQKDNFYQHNMEKLVNSQLGFKSPVEK-----GT 144  
 Db 122 ADTQSSSEDSTQNSAQDTPSDSKD---HHSDEADSRPEAGDSTQDSESEYRVGGGS 177  
 QY 145 QYLKQHTAMCKWQNEGPHSERLLESEPAVTLVPE---QFSNANVDQSSPKDDHSDTN 201  
 Db 178 EGSSSGDGS--EFDDGEG-----MQSDDPGSTRDRGHTRMSAGIRSESKDHEPTS 229  
 QY 202 SEESRDNQ--QFTVHKLAKAKQTMEDQ-----REARS-----HOKCKAKCHAP 246  
 Db 230 TQSDSDSDQVEFSRKSFRSRVSEEDRGELADSNRSQTQSVSTEDFRSEKSRSETQE 289  
 QY 247 ACAGCQOQETDVVSPLSDTSGEDVGV-----TGLKNANRLNQESSLGNSPPF 295  
 Db 290 DTATQSQSDSPGQDPSSSES-SEEAGPESQESSQEGVASEGRGNDPNTSQTGQDQ 348  
 QY 296 EKSESESPMDVNSKNSCQ-----DSEADETSPGFDEQ-----EDSSS-----AQTN 340  
 Db 349 DSSSEDRNTFSSSESQSTEQGSESNESLSLSEESQESAQDESSSQEGIQSQSAS 408  
 QY 341 KPSRFQPREADTELKRSKAKGGEIRLHFQEGESRAGMNDVNAKRPGETSS-----393  
 Db 409 RESRSQESQEQDSEENRDS-----DSQDSRSRSEESNS--TGSTSSSEEDNHP 457  
 QY 394 LNVECRNSK 402  
 Db 458 KNIEADNRK 466  
 RESULT 13  
 ID SR40 YEAST STANDARD; PRT; 406 AA.  
 AC P32583;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last annotation update)  
 DE Suppressor protein SRP40.  
 GN SRP40 OR YKR092C OR YKR412A.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 28383 / FL100;

Lalo D., Carles C., Sentenac A., Thuriaux P.;  
Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE FROM N.A.  
MEDLINE=94205265; PubMed=8154186;  
Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
"The complete sequence of a 15,820 bp segment of Saccharomyces  
cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three  
new open reading frames";  
RT Yeast 9:1349-1354 (1993).  
CC -|- FUNCTION: Not known; weak suppressor of a mutant of the  
CC subunit AC40 of DNA dependent RNA polymerase I and III.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L11275; AAA35091.1; -  
CC EMBL; J73541; CAA51946.1; -  
CC EMBL; Z28317; CAA82171.1; -  
CC PIR; S38170; S38170.  
CC GermOnline; 140071; -  
CC SGD; S0001800; SRP40.  
CC GO; GO:0005730; C:nucleolus; IDA.  
CC InterPro; IPR007718; SRP40\_C.  
CC Pfam; PF05022; SRP40\_C; 1.  
CC DOMAIN 25 314 ASP/SER-RICH  
CC CONFLICT 400 400 G -> N (IN REF. 1).  
CC SEQUENCE 406 AA; 14015 MW; 8EA007695AF4BALD CRC64;  
Query Match 3.0%; Score 154; DB 1; Length 406;  
Best Local Similarity 18.6%; Pred. No. 0.064;  
Matches 73; Conservative 73; Mismatches 174; Indels 72; Gaps 11;  
QY 40 LSKDAPVQFRVPPSSGCGALGRAGHRSATSLVFKQKITWMDTKGKTVESLSLHS 99  
DB 14 LSVKEIEIEKSS 73  
QY 100 KENNTEESMMSSVQDNFQHNWKEKLVNSQLGFDKSPVKGTVLYKHQHTAAWCK-W 158  
DB 74 ESS 133  
QY 159 QNEGPHSERLLESPPAVTLVPEQFNSNVDQSPKDPDHDSDTNSERDNNQFLTHVKLA 218  
DB 134 DNEAKETKAKTEPSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 179  
QY 219 NAKQTMEDQREARSHQKCKACHPAEACAGCQEQETDVVSPLSDTCSBVDGTLKN 278  
DB 180 SSS 221  
QY 279 ANRLNRQSSLCNPPPEKESESPMDVNS-----KNSQDSPEADETSPGFDEQSDS 334  
DB 222 -----DSS 272  
QY 335 SAQTANKPSRFQPPREADTELKRSAKGGEIRLHFQEGGESRAGMND--VNAKRPQSTS 392  
DB 273 SDSDSGSSSELETKEADESKAE-----ETPASSNESTFSSSSSSSAN 316  
QY 393 SLNVCEKNS--KQGRKDSKITDHPMVPKAE 422  
DB 317 KLNIPAGTDEIKGQK-----HFSRVDKSK 342  
RESULT 14  
IF38 HUMAN  
ID IF38 HUMAN STANDARD; PRT; 913 AA.  
AC Q99613; 000215;

DR EMBL; U91326; AAC27674.1; --  
DR EMBL; AK000739; BAA91352.1; --  
DR EMBL; BC001571; AAR01571.1; --  
DR Genew; HGNC:3279; EIF3S8.  
DR GK; Q99613; --  
DR MIM; 603916; --  
DR GO; GO:0005852; C:eukaryotic translation initiation factor 3 . . .; TAS.  
DR GO; GO:0003743; F:translation initiation factor activity; TAS.  
DR GO; GO:006446; P:regulation of translational initiation; TAS.  
DR InterPro; IPR008905; eIF3c\_N.  
DR InterPro; IPR000717; PCI.  
DR pfam; PF05470; eIF3c\_N; 1.  
DR SMART; SM00088; PCI; 1.  
DR Initiation factor; Protein biosynthesis.  
DR DOMAIN 164 189 ASP/GLU-RICH (ACIDIC).  
DR DOMAIN 164 189 POLY-GLU.  
DR DOMAIN 243 246 POLY-GLU.  
DR DOMAIN 291 294 POLY-GLU.  
DR CONFLICT 313 314 EK -> VR (IN REF. 2).  
DR SEQUENCE 913 AA; 105343 MW; CE5029F4EB51C1AA CRC64;  
Query Match 3.0%; Score 153; DB 1; Length 913;  
Best local Similarity 18.8%; Pred. No. 0.21; 321; Indels 312; Gaps 43;  
Matches 175; Conservative 125; Mismatches 321; Indels 312; Gaps 43;  
QY 198 SDTNSERSRDNQQLTHVKLAN-AKQTM---EDRQ----- 228  
DB 9 SDESSESLGSEELVTPVGVNGYKQPLLSDEBEDTKRVRSADKRPBELTLNLTIR 68  
QY 229 -GREARSHQK-----GKACHPAECA---GCOQETDVVS---RSLSDTGSVDYGTG 275  
DB 69 NAMKIRDVTCKLEBFELLGKAYGAKSIVDKGVPRFVIRILADLEYLNELWEDKGGK 128  
QY 276 LKANRLN-----ROESLGNPPFEK-----ESEPSPMDVNSKSCQSDSEADETSP 325  
DB 129 KMKNNNAKALSTLRQIKRKNR-DFESHITYKQNPQSDADEAKN-----EEDSGSS 182  
QY 326 GFDEQSDSSAQTAKPSRFQPREADTELKRSKAGGEIR--LHFOPEGGESRAGMNDV 383  
DB 183 DEDEDEGVSAATF-----LKKSEAPSGESRKLKQWDEDEDESDSD 228  
QY 384 NAKRPGSTSSLVNCRNSKQHRKDKSKITDHMRVP-----KAEKPKQCEKMHQRT 436  
DB 229 EDWDTGSTSS---DSDSEEBGKOTALASRFLKKAAPTDEDKKAEKREDKAKKCHDRK 285  
QY 437 ERKTPKTPPLSPDKKWLGTPIEMRMWRPGIRLPLRP-----SANHTVTIRVDL 489  
DB 286 SKRL-----DEEBEDNEGGEWERV-RGGVPLVKEKPMFAKGTETITHAVVIK--- 331  
QY 490 LRIGVPPKFPFTHFKDLNDKNHVKMPCSEONLYPVEDENGERRAAGSRWELIQTALLNRLT 549  
DB 332 -KLNEL-----LQARGKKGTDRAA--QIELLQ--LLVQIA 361  
QY 550 RPQNLKDAI---LKNVAYSKWDFTALI-----DFWDKYLEEAEAOHLQSIPLDMVK 600  
DB 362 AENNLGEGVIVKIFENIIAS-LYDYNPNLATYMKPEMKGCLD-----CINELMD 410  
QY 601 TALCLPNI-----CTQPIPL-----LQKKNHSLTM-----SQBQI 631  
DB 411 ILFANPFIVGNILFESENHNAOPLRVRCILTLVERMDDEETKINQNTDPSHQEV 470  
QY 632 ASILANAFCTPPRRNAKMSYSSYPDINRFLF-----EG 668  
DB 471 EHLKDEAQVCAIERVQRYLEEGTTEEV--CRIVLLRILHTYKPYDKAHQRLTPPEG 528  
QY 669 RSRKPEKLUK-----LFCYFRRTVEKPTGLVFTTQSLDEDFPFWERCEKL 715  
DB 529 SSKSEQDAENEGEDSAVLMERLCKYIYAKDRTDIRTCAILCHIVHHAHLSRWQARDL 588  
QY 716 LTELHVTYEGTIEGNGQGLQVDFAKRFVGGVTS--AGLVQIEFRLINPELLIVSRFLT 773  
DB 589 MLMSHL--QDNIO-HADPPVQIILYNTVQLGICAFRQSLTKDA----- 629

QY 774 EVLDHNECLIIITGEYSEYTGTYASTYRWARSHEDESRDDMQRRTEITVAIDALHFRRY 833  
DB 630 ----FNALLDIQSSRAKELLGGLLSRLSLQBRNOEQEKVRRRQ----- 670  
QY 834 LQDFVPEKIRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQIL--AA 891  
DB 671 ----VFFHLHINLELLECVYL-----VSAMLLLEIPYMAAHESDARRRMISKQFHQL 718  
QY 892 AVAERDVVYFTFGDSBELMR-----DIVSMGTFLTERKUTVGEVYKL----- 932  
DB 719 RVGERQPL--LGPPESEEHVVAASKAMKMGDKWTKCHSFINEKKN--GKVDLFPFEADK 774  
QY 933 ----LLRYNNECRNCSTPGPDIKLYPFYHAV 961  
DB 775 VRTMLVRKIQEE-----SLRTLYFTYSSV 798  
RESULT 15  
DPOZ MOUSE  
ID DPOZ\_MOUSE STANDARD; PRT; 3122 AA.  
AC Q61493; Q9JMD6; Q9QW6;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE DNA polymerase zeta catalytic subunit (EC 2.7.7.7) (Seizure-related protein 4).  
GN REV3L OR POLZ OR SEZ4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID:10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/Ola; TISSUE=Testis;  
RX MEDLINE=99202265; PubMed=10102037;  
RA Van Sloun P.P.H., Romeijn R.J., Eeken J.C.J.;  
RT "Molecular cloning, expression and chromosomal localisation of the mouse Rev3l gene, encoding the catalytic subunit of polymerase zeta.";  
RL Mutat. Res. 433:109-116 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kajiwara K.;  
RT "Molecular analyses of Sez4 encoding murine homologue of yeast REV3 in brain neurons.";  
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 2368-3122 FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Embryonic brain;  
RX MEDLINE=96216731; PubMed=8645260;  
RA Kajiwara K., Nagawawa H., Shimizu-Nishikawa K., Ookura T., Kimura M., Sugaya E.;  
RT "Molecular characterization of seizure-related genes isolated by differential screening.";  
RL Biochem. Biophys. Res. Commun. 219:795-799 (1996).  
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + [DNA] (N).  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- SIMILARITY: Belongs to the DNA polymerase type-B family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF083464; AAC98785.1; --  
DR EMBL; AB031049; BAA90768.1; --  
DR EMBL; D78644; BAAL1461.1; --  
DR PIR; T17202; T17202.  
DR MGD; MGI:1337131; Rev31.  
DR InterPro; IPR006172; DNA\_pol\_B.

DR InterPro; IPR006134; DNA\_pol\_B\_dom.  
 DR InterPro; IPR006133; DNA\_pol\_B\_exo.  
 DR InterPro; IPR004578; Pol2.  
 DR Pfam; PF00136; DNA\_pol\_B; 1.  
 DR Pfam; PF01104; DNA\_pol\_B\_exo; 1.  
 DR PRINTS; PR00106; DNAPOB.  
 DR SMART; SM00486; POLBc; 1.  
 DR TIGRFAMs; TIGR00592; pol2; 1.  
 DR PROSITE; PS00116; DNA\_POLYMERASE\_B; 1.  
 KW Transferase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; DNA repair; Nuclear protein; Zinc-finger.  
 FT ZN\_FING 3034 3049  
 FT C4-TYPE (POTENTIAL).  
 FT ZN\_FING 3078 3096  
 FT C4-TYPE (POTENTIAL).  
 FT CONFLICT 92 92  
 FT A -> A (IN REF. 2).  
 FT CONFLICT 294 294  
 FT A -> T (IN REF. 2).  
 FT CONFLICT 578 578  
 FT E -> Q (IN REF. 2).  
 FT CONFLICT 609 609  
 FT R -> Q (IN REF. 2).  
 FT CONFLICT 1278 1278  
 FT L -> P (IN REF. 2).  
 FT CONFLICT 1298 1298  
 FT L -> F (IN REF. 2).  
 FT CONFLICT 1416 1416  
 FT P -> L (IN REF. 2).  
 FT CONFLICT 1848 1848  
 FT A -> T (IN REF. 2).  
 FT CONFLICT 2368 2368  
 FT V -> G (IN REF. 3).  
 SQ SEQUENCE 3122 AA; 350654 MW; A39846CAF7365BA8 CRC64;

Query Match 3.0%; Score 153; DB 1; Length 3122;  
 Best Local Similarity 19.4%; Pred. No. 1;  
 Matches 217; Conservative 148; Mismatches 387; Indels 368; Gaps 56;

QY 2 SAGPGCECTKPRW---DAAATSPPAASDARSFGRORRVLDSDKAPVQFRVPSSSGC 58  
 DB 1122 SSPRC-----WSPTDPAEETMAAAEKESMLFKGPNVNTK--TVSPRVGKASAR 1171

QY 59 ALRGAQHRSATSLVF---KQKITSMQDTKGIKTVESESLHSEKNNTHREEMSSV 114  
 DB 1172 AQVKKSKARLANSVYTNKRNQNTTLVDGKKKPRAKQORANEKLSLKHAIAD- 1230

QY 115 QKDNFYQHMEKL---ENVSQLGFDKSPVEKGTQVLKQHQTAAMCKWNEGFHSE-LL 169  
 DB 1231 --EKMKPHSEALTFNHQVSEL-----TSSGAQALSKQK-----EWSQTGPAVDHPLL 1278

QY 170 ESEPPAVTLVPEQFNA-----NVD-----QSSPKDDHSDT-----NSEE 204  
 DB 1279 PAQPTGIS-AQRLNCFSSLESKSVLDRTFPPSRDSDSHSVYSSIGPGISKINIQ 1337

QY 205 SHDNQOFLTHVLAKAKOTMEDEQREARSHQCKGACACAPACACQCOQEEEDVVSSEPL 264  
 DB 1338 SHNQAMFTRKETTLTQKSIIFDLNHLQVAQS-----TQVCSGIIISPKTE-ESSSTQ 1389

QY 265 SDTGSVDGTGLKNANRLNRQESSLGNSSPPFEKESPEFMDVDNSKNS---CQDSEADE 321  
 DB 1390 KNGCS-----SMGKLNYSLSLESKP--EQVCAPNFLHCKDSQQQTVSVSEQSKTSE 1439

QY 322 ETPSGDEQEDSSAQAN-----QTPNCFVTSLSKPIQIAWEQKQRFILDMNSFNFKPKYKQRLSE 1496  
 DB 1440 TCSPGNAASES---QTPNCFVTSLSKPIQIAWEQKQRFILDMNSFNFKPKYKQRLSE 1496

QY 350 ADTELKRSAGGEIRLHFQF--EGGESRAGMNDVNAKRGSTSSLVN--ECRNSKQHGK 406  
 DB 1497 AISQTKALSQCKNQNVSTPSVFGEGSGLAVLKELQKQKQAQSTNVVDSTSTHQPK 1556

QY 407 KDSKITDHFMRVPAEDKREKQCEMKHQRTERKI--PKYIPPH-----KPSRFQPR--E 349  
 DB 1557 NISVSNEH-----XKANKRTPTVTSRKPRTPTKPKQTPRLKVDPLNLTSGHLDS 1612

QY 448 LSPDKXWL-GTPIEEMRMPCGIRLPFLRPSANHTVTRVLLRIGE----- 494  
 DB 1613 LSDSPILFSDPGFE-----SCYSLEDSLSPENY---NFDINTIGTGFCSFYSGSQF 1663

QY 495 VP--KPPPTHF-----KDLWKN-----HVKMPCEQ-----NLYP- 523  
 DB 1664 VPADONLPQKFLSDAVQDLFPQGAIDKSELLSHDROSCSEKHHVSDSSWIRASTLNPE 1723

QY 524 ----VEDENGERRAAGSRWE-----LIQTALLNR----- 547

DB 1724 LFEKVAMDNNEHRHSQWKNSEFHLTSHSNIMESFCVQQAENCLTKSRNLNRSSVSKEV 1783  
 QY 548 -LIREPO-NLKDAILKYN-----VAYSKWDFTALIDFWDKVLEBAQAQ--HLYQSILPD 597  
 DB 1784 FLSLFOANSSDWIQGHNRKEADQSLHSANTSTFTILSSPDGELVDAASEDELEYVSRND 1843  
 QY 598 MVKIAL-CLPNICTQPIPLLKQKNHSIT-----MS-----QBOIASLLANAF--- 639  
 DB 1844 VLITAPDSSPRSTSSPL---QSKNGSFTPTAHILKPLMSPSPSREETVATLLDHLSEA 1899  
 QY 640 -----PCTFFPRNAKMSSEYSSYPDINFENRLFEGSRSSKPEKLTLCFYFRRVTEKKPTG 694  
 DB 1900 IYQEPFCSNP-----SDVPEKPREIGGRLLMVETRLND 1933  
 QY 695 LVTFTRO-SLEDDFPEWERCEKLLTFLHVTYBGTIEGNGQGMLOVDFAHRFYGGVTSAGL 753  
 DB 1934 LIRFEGDFSLEGLRWKTAFSAMTQ--NPRGSPLRNGQAVVNKESN-----SHKM 1983  
 QY 754 VQEBIRFLINPELIVSRLFTIEVDHNECLIIITGTEQYSEYTGTAETRWARSHEDRSERD 813  
 DB 1984 VEDK-KIVIMP-----C-----KYAPSRQLVQAWLQAKEEYERSKK- 2018  
 QY 814 DWQRTTEIVAIDALHFRYLDQFVPEKIRRELANKAYCGFLRPGVS----- 859  
 DB 2019 --LPKTELTPVK-----SAENVSPSLNPGDTCVSPQVDKCPHTLSSSAHTKE 2065  
 QY 860 --SENLSAVATGNWCGGAFGAGDARLKALIQILAAVAERD 897  
 DB 2066 EVSKSQIALQSTTGCs-----QTLAAASAAPVEED 2097

Search completed: May 26, 2004, 18:41:04  
 Job time : 24.6533 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:28:37 ; Search time 173.272 Seconds  
(without alignments)  
13270.296 Million cell updates/sec

Title: US-09-302-812-3  
Perfect score: 7208  
Sequence: 1 ggcgtctgggaagtggag.....agaaaaaaaaaaaaaa 4069

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cpn2.1/uspto.spool.p/6333148/runat.26052004.150052.5654/app.query.fasta_1.12437
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=6333148 @CGN 1.1.414 @runat.26052004.150052.5654 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database:

1:	Geneseq1980s.*
2:	Geneseq1990s.*
3:	Geneseq2000s.*
4:	Geneseq2001s.*
5:	Geneseq2002s.*
6:	Geneseq2003as.*
7:	Geneseq2003bs.*
8:	Geneseq2004s.*

Result No.	Score	Query Match	Length	DB ID	Description
1	5190	72.0	976	5	Aae25630 Human pol
2	5190	72.0	976	5	Aau76021 Human pol
3	5190	72.0	976	5	Aau76012 Human pol
4	5190	72.0	976	6	Abg72279 Human pol
5	4635.5	64.3	977	5	Aae25629 Bovine po
6	4635.5	64.3	977	5	Aau76020 Bovine po
7	4635.5	64.3	977	5	Aau75799 Bovine po
8	4635.5	64.3	977	6	Abg72278 Bovine po
9	4395	61.0	968	5	Aae25631 Murine po
10	4395	61.0	968	5	Aau76022 Mouse pol

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

XX	AAE25630	standard; protein; 976 AA.
XX	AAE25630	
XX	AAE25630	
DT	04-NOV-2002	(first entry)
DE	Human poly adenosine diphosphate-ribose glycohydrolase (PARG).	
XX	Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme; ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic; cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity; Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiact; cytoskeletal; vasotrophic; neuroprotective; anticonvulsant; gene therapy; antisense therapy.	
XX	Homo sapiens.	
XX	US6395543-B1.	
XX	28-MAY-2002.	
PF	23-FEB-2000;	2000US-00511507.
XX	01-MAY-1998;	98US-0083768P.
XX	30-APR-1999;	99US-00302812.
PA	(KENT ) UNIV KENTUCKY RES FOUND.	
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;	
XX	WPI; 2002-535641/57.	
DR	N-PSDB; AAD42082.	

# ALIGNMENTS

RESULT 1  
AAE25630  
ID AAE25630 standard; protein; 976 AA.

XX	AAE25630		11	4395	61.0	968	5	AAU76013
XX	AAE25630		12	4395	61.0	968	6	ABG72280
XX	AAE25630		13	1081	15.0	768	4	ABB59491
XX	AAE25630		14	1081	15.0	768	5	AAE25632
XX	AAE25630		15	1081	15.0	768	5	AAU76023
XX	AAE25630		16	1081	15.0	768	5	AAU76014
XX	AAE25630		17	1081	15.0	768	6	ABG72281
XX	AAE25630		18	463	6.4	726	5	AAE25633
XX	AAE25630		19	463	6.4	726	5	AAU76024
XX	AAE25630		20	463	6.4	726	5	AAU76015
XX	AAE25630		21	463	6.4	726	6	ABG72282
XX	AAE25630		22	344.5	4.8	819	4	ABG20721
XX	AAE25630		23	210	2.9	100	4	ABG11103
XX	AAE25630		24	178	2.5	33	5	AAE25652
XX	AAE25630		25	177	2.5	1163	4	AAU37362
XX	AAE25630		26	175	2.4	1111	4	AAU37362
XX	AAE25630		27	175	2.4	1274	6	ABU18980
XX	AAE25630		28	175	2.4	1274	6	ABM71209
XX	AAE25630		29	174	2.4	1274	6	ABU15913
XX	AAE25630		30	174	2.4	4873	6	ABO14747
XX	AAE25630		31	173.5	2.4	748	4	AAU58340
XX	AAE25630		32	173	2.4	1948	5	ABP73774
XX	AAE25630		33	170.5	2.4	768	4	AAU78686
XX	AAE25630		34	170.5	2.4	2703	4	ABB60074
XX	AAE25630		35	170	2.4	335	4	ABG09631
XX	AAE25630		36	169.5	2.4	1803	4	ABB65391
XX	AAE25630		37	169	2.3	1294	6	ABP75935
XX	AAE25630		38	168.5	2.3	2194	4	AAU40114
XX	AAE25630		39	168	2.3	1270	7	ADC37572
XX	AAE25630		40	167	2.3	31	5	AAE25634
XX	AAE25630		41	167	2.3	31	5	AAE25651
XX	AAE25630		42	167	2.3	31	5	AAU76025
XX	AAE25630		43	167	2.3	31	5	AAU76016
XX	AAE25630		44	167	2.3	31	6	ABG72283
XX	AAE25630		45	166	2.3	1163	6	ABU62602







Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	72.00%	Indels:	0
DB:	5	Gaps:	0
US-09-302-812-3 (1-4069) x AAU76021 (1-976)			
QY	167	ATGAATCGCGGCCCCCGCTGTGTGAACCTCTGACCAAAAGCGACCCGCTGGGCGCGGTACA	226
DB	1	MetAenAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr	20
QY	227	ACTTCGCGCTGCTTCGACGCGCGGAGCTTCCGACGAGGCGAGGCGGCTTCGAC	286
DB	21	ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp	40
QY	287	CCCAAGACGCTCACGTGCAGTTCAGGGTCCCAACCGCTCTCGCAGCGCTCGCTCCAGGG	346
DB	41	ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly	60
QY	347	CAGGCGGACAGCACAGAGCGCGCCACTCGCTGCTGTTTCAAAACAAAGACTATTACC	406
DB	61	GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr	80
QY	407	AGTTGGATGCACACTAAAGCAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA	466
DB	81	SerTrpMetAspThrLysGlyLysThrAlaGluSerGluSerLeuAspSerLysGlu	100
QY	467	AACAACAATACAAAGATAGATCAATCATGATGAGTTCGTACAAAAGATACTTTTACC	526
DB	101	AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTrpGln	120
QY	527	CATAATGTGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAGTCACTCACTGAA	586
DB	121	HisAsnValGlnLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu	140
QY	587	AAAAGTACACAGATTTTGAACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG	646
DB	141	LysSerThrGlnTrpLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu	160
QY	647	GGAAAAACACGAGAGAGCTTTTGGAAAGTGAACCTCAACAGTCAACCTCGGTACCAAG	706
DB	161	GlyLysHisThrGlnGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu	180
QY	707	CAGTTTATGATGATACATGATGCTGCTCACTCACTCACTCACTCACTCACTCACTCACT	766
DB	181	GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp	200
QY	767	AGTGAAGAGATAGACAAATCAACAGTTCCTCAACTGTAAAGCTTGCAAAATGCAAG	826
DB	201	SerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAsnAlaLys	220
QY	827	CAGACTACGGAAGATGAACACGCGCAGAGAACCCAAAGCCACAGAGTGCAGCAAGTCT	886
DB	221	GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer	240
QY	887	TCCCATCTCGGAGAGCTGCAAGTGTCAAGATGATGATGATGATGATGATGATGATGATG	946
DB	241	CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLysLeuValProLys	260
QY	947	AGTCCATGTGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAACA	1006
DB	261	SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys	280
QY	1007	TTGATTAGCAAGAAAGTTGCTAGGAAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1066
DB	281	LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro	300
QY	1067	GAATCACCAGTGCATGATTAATCTAAATAGTCTCAAGACTCAGAAGCAGATGAG	1126
DB	301	GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu	320
QY	1127	GAGACAAGTCCAGGTTTGTATGAAACAAGATGTTGTTCTCTCTCTCTCTCTCTCTCTCT	1186
DB			
DB	321	GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys	340
QY	1187	CCTTCAAGGTTCCAAAGACGAGACCTGACATTTAGTAATTTAGGAAACGGTACTCTAAG	1246
DB	341	ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTrpSerThrLys	360
QY	1247	GCGCGTGAAGTTAGATTACATTTTCAATTTGAAGGAGGAGAGTGCCTGCAATGAAT	1306
DB	361	GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn	380
QY	1307	GATTAAATGCTAAACTACCTCGAAATATTCTAGCTGAAATGTAGAAATGCAGAAATCT	1366
DB	381	AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer	400
QY	1367	RAGCAACATGGAAGAAAGGATTTCTAAATCACAGATCATTTGATGAGCTGCCCAAGCA	1426
DB	401	LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla	420
QY	1427	GAGGACAGAGAAAGAACACAGTGGGAAACCAACATCAAGAACAGAAAGAGATCCCT	1486
DB	421	GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro	440
QY	1487	AAATAGCTCCACCTCACCTTTCTCCAGATAGAGTGGCTTGGACTCCCATTTAGGAG	1546
DB	441	LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu	460
QY	1547	ATCAGAGAAGATCGCTCGGTGGGATCCGCTCGCTCTCTGAGACCATCTGCCAATCAC	1606
DB	461	MetArgArgMetProArgCysGlyLysLeuLeuProLeuLeuArgProSerAlaAsnHis	480
QY	1607	ACAGTAACTATTTCGGGTAGATCTTTTCGAGCAGGAGAGTTCCTAAACCTTTTCAACA	1666
DB	481	ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr	500
QY	1667	CATTATAAGATTGTTGGGATTAACAGCATGTTAAATGCTTGTTCAGAACAAAAATTG	1726
DB	501	HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu	520
QY	1727	TACCCAGTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1786
DB	521	TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln	540
QY	1787	ACTGCATCTTCTCAACAAATTTACACGACCCCAAACTTGAAGAGTGTCTATTCTGAAT	1846
DB	541	ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr	560
QY	1847	AATGTGCATATTCTAAGAAATCGGACTTTTACAGCTTTGATCGATTTCTCGGATAAGG	1906
DB	561	AsnValAlaTyrSerLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal	580
QY	1907	CTTGAAGAAGCAGAGCTCAACATTTATATCATCTTATCATCTTATCATCTTATCATCT	1966
DB	581	LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle	600
QY	1967	GCATCTGTCTGCAAAATATTTCACCCAGCCAAATACCATCTCTCTGAAACAGAGATCA	2026
DB	601	AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn	620
QY	2027	CATTCCATCAATGTCGAGGAAACAGTTCGAGTCTTTTAGCTAATGCTTCTTCTCTGC	2086
DB	621	HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys	640
QY	2087	ACATTTCCACGAGAAATGCTAAGATGAAATCGAGTATTCTAGTATCCAGACATTAAC	2146
DB	641	ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTrpProAspIleAsn	660
QY	2147	TTCAATCGATTGTTGAGGACGTTTCATCAGGAAACCGGAGAACTTAAACGCTCTTC	2206
DB	661	PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe	680
QY	2207	TGCTACTTTTGAAGAGTACAGAGAAAAAACCCTACTGCGTGGTGGTGCATTTTACAG	2266
DB	681	CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln	700



QY 2267 AGCTTGAAGATTTCAGAGATGGAGAGATGTGAACACCCCTTGACACGATTCGATGTC 2326  
 DB |||||  
 QY 701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720  
 DB |||||  
 QY 2327 ACTTACGAAGGTACCATAGAGAAATGCGCAAGGCATGCTACAGGTGGATTTTGCRAAT 2386  
 DB |||||  
 QY 721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740  
 DB |||||  
 QY 2387 CGTTTTGTGGAGGTGTAAACAGTCAGGACTTGTGCAAGAGAAATCGCTTTTAA 2446  
 DB |||||  
 QY 741 ArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPheLeu 760  
 DB |||||  
 QY 2447 ATCAATCCGTGAGTTTACATTCACGGCTCTTCACTGAGGTGCTGGATCAACAATGAATGT 2506  
 DB |||||  
 QY 761 IleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
 DB |||||  
 QY 2507 CTAATTATTCAGAGTACTGACAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566  
 DB |||||  
 QY 781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800  
 DB |||||  
 QY 2567 TGGTCCCGAGCCAGAGATGGAGTGAAGGAGCAGCTGGAGCGCGCTGCACTGAG 2626  
 DB |||||  
 QY 801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820  
 DB |||||  
 QY 2627 ATCGTTGCATCGATGCTTCACTTACAGCGCTACCTCGATCAGTCTTGTGCTGAGAAA 2686  
 DB |||||  
 QY 821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840  
 DB |||||  
 QY 2687 ATGACGCGAGCTCAACAGGCTTACTGTGTGATTTCTCCGTCTGGAGTTTCTTCAGAG 2746  
 DB |||||  
 QY 841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860  
 DB |||||  
 QY 2747 AATCTTCTGACGTGCGCACAGAACTCGGGCTGTGTGCTGCTTGGGGTGCATCCAGG 2806  
 DB |||||  
 QY 861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880  
 DB |||||  
 QY 2807 TTAAAGCCTTAATACAGATATGTCAGCTGCTGAGCTGAGCAGAGATGTGTTTATTTTC 2866  
 DB |||||  
 QY 881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyrPhe 900  
 DB |||||  
 QY 2867 ACCTTTGGGGACTCAGATTGATGAGACATTTACAGCATGACATTTTCTTACTGAA 2926  
 DB |||||  
 QY 901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920  
 DB |||||  
 QY 2927 AGGAACCTCACTGTGGAGATGTATAGCTGTGCTACGATCTACATGACATGACATGC 2986  
 DB |||||  
 QY 921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluCys 940  
 DB |||||  
 QY 2987 AGAAACTGTTCACCCCTGGACACAGATCAAGCTTTATCCATTATCATATACCATGCTGC 3046  
 DB |||||  
 QY 941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960  
 DB |||||  
 QY 3047 GAGTCTCTGACAGACCGCTGACCATTCAGGCGCAAGAGACAGGACC 3094  
 DB |||||  
 QY 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
 DB |||||

RESULT 3  
 AAU76012

ID AAU76012 standard; protein; 976 AA.

XX AAU76012;

XX AAU76012; (first entry)

DT 08-MAY-2002 (first entry)

XX Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 DE Human poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.

OS Homo sapiens.  
 XX US6333148-B1.  
 PN 25-DEC-2001.  
 PD 30-APR-1999; 99US-00302812.  
 PF 01-MAY-1998; 98US-0083768P.  
 PR (KENT) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-153820/20.  
 DR N-PSDB; ABK14494.  
 XX Screening compounds for modulation of poly (ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX Claim 3; Col 55-60; 80pp; English.  
 PS The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly (ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the human PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX Sequence 976 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 976  
 Score: 5190.00 Matches: 976  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 72.00% Indels: 0  
 DB: 5 Gaps: 0

US-09-302-812-3 (1-4069) x AAU76012 (1-976)

QY 167 ATGAATCGGGCCCCCGCTGTGAACCCCTGCACCAAGCGACCCGCTGGGCGCCGTACA 226  
 DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
 QY 227 ACTTCGCGCGTGTTCGAGCCCGGAGCTTCCGACGACGACGAGCGGCTCTCCGAC 286  
 DB 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp 40  
 QY 287 CCCAAGGACGCTCAGCTGCTGAGTTCAGGTCCACCGTCTCCACAGCTCTCCGACGG 346  
 DB 41 ProllysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60  
 QY 347 CAGCGGACAGCAGCAGAGCGCCACCTCGCTGTGTTTTCACAAAGACATATTACC 406  
 DB 61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80  
 QY 407 AGTTGGATGGACACTAAAGGAATCAGACGCGGATCAGAAAGTTTGTATAGTAAAGAA 466  
 DB 81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100  
 QY 467 AACCAACAATCAAGAATAGATTCATGATGATGCTGTGTACAAAAAGATAACTTTTACCAA 526  
 DB 101 AsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyrGln 120

527 CAPAATGTGAAAAATTAGTAATGTTTCTCAGCTAAGTCTTGATAGTCACTCACTGAA 586  
121 HieasnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140  
587 AAAAGTACACAGTATTTGAACACGACATCAGCTGCGCAATGTAGTGGCAAAATGAA 646  
141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTyrGlnAsnGlu 160  
647 GGGAAACACACGGAGCAGCTTTTGGAAAGTGAACCTCAAAACAGTAACCCCTGGTACCAGAG 706  
161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180  
707 CAGTTTAGTAATGCTAAATGATCGGTACCTCAAAATGATGATGATGATGATGATGATGAT 766  
181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200  
767 AGTGAACAGAAATAGAGCAATCAACAGTCTTCAACACTGTCAAACTGTCAAACTGTCAAA 826  
201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220  
827 CAGACTACGGAAGATGAACACCGCAGAGAAGCCAAAGCCACCAAGAGTGCAGCAATGCT 886  
221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240  
887 TGCATCTCTGGGAGAGCTGCGAAGTGTGCAAGATGATGATGATGATGATGATGATGATGAT 946  
241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260  
947 AGTCCATTGTGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAATGATGACACAAA 1006  
261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280  
1007 TTGATTAGACAAAGTGTGCTAGGAAATTCCTCCATTTGAGAGAGGAAGTGAACCC 1066  
281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGluPro 300  
1067 GAATCACCGATGGATGGTAATCTTAATAGTGTCTCAAGACTCAGAGCAGAGATGAG 1126  
301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320  
1127 GAGCAAGTCCAGGTTTGTATGAACAAGATGTTAGTCTCCTCCCAACACGCAAAATAAA 1186  
321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340  
1187 CCTTCAAGGTTCCAAGCAGAGACGCTGACATGTAATTTAGAAACGGTACTCTACTAAG 1246  
341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360  
1247 GCGGTGAGTGTAGTTACATTTCCAAATTTCAAGGAGGAGAGTCCGACCTGGATGAAT 1306  
361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380  
1307 GATTAAATGCTAACTACCTCGAAATATTTCTAGCTGAATGTAGAAATGCAAGAAATCT 1366  
381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400  
1367 AAGCAACATGGAATAAGGATTTCTAAATTCAGATTCAGATTCATTTGATGAGACTGCCCAAGCA 1426  
401 LysGlnHisGlyLysLeuAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420  
1427 GAGGACAGAAAGAAACAGTGGGAAACCAACATCAAGAACAGAAAGGAAGATCCCT 1486  
421 GluAspArgArgLysGluGlnThrGluThrLysHisGlnArgThrGluArgLysIlePro 440  
1487 AAATACGTTCCACCTCACCTTTCTCCAGATAGAAAGTGGCTGGAACTCCCATGGAGAG 1546  
441 LysTyrValProProHisLeuSerProAspLysLysTyrLeuGlyThrProIleLeuGlu 460  
1547 ATGAGAAGATGCCCTGGGTGGGATCCGGCTGCTCTCTTGAGACCATCTGCCAATCAC 1606  
461 MetArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480  
1607 ACAGTAACATTTCCGGTAGATCTTTTTCGCGAGCAGGAGGATTTCTTAAACCTTTTCCAACA 1666

481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500  
1667 CATTTATAAGATTTGTGGGATAACAAGCATGTTTAAAAATGCTTTGTTCAGAACAAAAATTG 1726  
501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520  
1727 TACCAGTGGAGATGAGATGCTGAGCGAAGTCCGCGGAGCCGGTGGGAGCTCATTCAG 1786  
521 TyrProValGluAspGluAsnGlyIleArgThrAlaGlySerArgTyrGluLeuIleGln 540  
1787 ACTGCATCTCTCAACAAATTTTACAGACCCCAAACTTGAAGCATGCTATCTTGAATAC 1846  
541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560  
1847 AATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTTGTATTCGATTTCTGGGTAAGGTA 1906  
561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580  
1907 CTTTGAAGACGACAGCTCAACATTTATATCAGTCCATCTTGCCTGATATATGGTCAAAATT 1966  
581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600  
1967 GCATCTCTGCTGCCAAATATTTCACCCAGCCCAATACCATCTCTGAAACAGAGATGAAT 2026  
601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620  
2027 CATTCATCATGATGTCGAGGACAGATTCAGCTCTTTTGTAGTATGCTTCTTCTCTGC 2086  
621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640  
2087 ACATTTCCAGCAGAAATGCTAAGATGAAATCGAGATTTCTAGTTACCCAGACATTAAC 2146  
641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660  
2147 TTCATTCGATTTGTTGAGCAGCTTCATCAAGAAACCCGAGAGAAACCTTAAACCCCTCTC 2206  
661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGlyLysLeuLysThrLeuPhe 680  
2207 TGCTACTTTTGAAGAGTCAAGAGAAACCTTACTGCTGCTGCTGATTTACATTTACAGACAG 2266  
681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700  
2267 AGTCTTCAAGATTTCCAGAAATGGAAGATGTGAAACCCCTTGACACCATGTCATGTC 2326  
701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720  
2327 ACTTACGAAGGTACCATAGAAATAATGGCCAGGCATGCTACAGGTGGATTTTGCAAT 2386  
721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740  
2387 CGTTTGTGAGGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAATCCGCTTTTAA 2446  
741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPheLeu 760  
2447 ATCAATCTCAGTGTATTATTTCACGGCTCTTCTCAGGTGCTGGATCAATGAATGT 2506  
761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
2507 CTAAATTATCAGGTACTGAGCAGTACAGTGAATACAGGCTATGCTGAGACATATCGT 2566  
781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800  
2567 TGGTCCCGGAGCCACGAGATGGAGTGAAGGACGACCTGCGAGCGCGCTGCACCTGAG 2626  
801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820  
2627 ATCGTGTGCATGCTCTTCTTCACTTCAGACGCTACCTCCGATCATGTTGTGCTCGAGAAA 2686  
821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840  
2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGGATTTCTCCGCTCTGAGGTTTCTTCAGAG 2746



827 CAGACTACGAGATGAACACGCGCAGAGAGCCAAAGCCACAGAGTGCAGCAAGTCT 886  
|||  
221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240  
|||  
887 TGCATCTCTGGGGAAGACTGTGCAAGTTGTGACCAAGATGAGTAGACGTGTGCGCAAG 946  
|||  
241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuValProLys 260  
|||  
947 AGTCCATTGTTCAGATGTGGCTCTGAGGATGTGGTACTGGGTCAAAAAATGACACAAA 1006  
|||  
261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280  
|||  
1007 TTGATTAGACAAAGAAAGTTGCCCTAGGAAATTCCTCCATTGAGAGGAAAGTGAACCC 1066  
|||  
281 LeuLeuArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300  
|||  
1067 GAATCACCAGTATGATGGATTAATCTAAATAAGTTGTCAAGACCTCAGAGGAGATGAG 1126  
|||  
301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320  
|||  
1127 GAGACAAGTCCAGCTTTTGATGAACAAGAGATGGTAGTTCTCCCAAAACAGCAAAATAA 1186  
|||  
321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340  
|||  
1187 CCTTCAAGGTTCCAAGCAGAGACGCTGCATTTGAATTTAGGAAACCGTACTCTACTAAG 1246  
|||  
341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360  
|||  
1247 GCGGTGAAGTATGATTTACATTTCCATTTGAGGAGGAGAGTCCGACTCGAATGAAT 1306  
|||  
361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380  
|||  
1307 GATTTAATGCTAACTACCTGGAATAATTTCTAGCTGATGAGTGTGAGTGTGAGAAATCT 1366  
|||  
381 AspLeuAsnAlaLysLeuProGlyAsnLysSerSerLeuAsnValGluCysArgAsnSer 400  
|||  
1367 AAGCAACATCGAAAGAAAGGATTTCAAATCAAGATCATTTGATGAGACTGCCCAAGCA 1426  
|||  
401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420  
|||  
1427 GAGACAGAGAAAGACAGTGGGAACCAACATCAAGAACAGAAAGAGATCCCT 1486  
|||  
421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440  
|||  
1487 AAATACGTTCCACCTCACCTTCTCCAGATAAGAAGTGGCTTGAACCTCCCATTTGAGGAG 1546  
|||  
441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460  
|||  
1547 ATGAGAGAATGCTCGGTGTGGATCCGGTGTCTCTTGAGACCATCTGCCAATCAC 1606  
|||  
461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480  
|||  
1607 ACAGTAATATTCCGTTAGATCTTTGCGAGCAGAGAGTTCCTAAACCTTTTCCAACA 1666  
|||  
481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500  
|||  
1667 CATTATAAGATTTGTGGGATAACAAGCATGTTAAATGCTTGTTCAGAACAAATTTG 1726  
|||  
501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520  
|||  
1727 TACCCAGTGGAGATGAGATGGTGGAGCAACTCGGGGAGCCGTGGGAGCTCATTTGAG 1786  
|||  
521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540  
|||  
1787 ACTGACTCTCAACAAATTTACAGACCCCAAACTTGAAGGATGCTATTCTGAAATAC 1846  
|||  
541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560  
|||  
1847 AATGGGCATATTTCAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAGGTA 1906  
|||  
561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580  
|||  
1907 CTTGAAGAGCAGAGAGCTCAACATTTATACAGTCCATCTTGCCTGATATGCTGAAATTT 1966  
|||

581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600  
|||  
1967 GCACCTCTGTCTGCCAAATATTTGCACACCAGCCAAATACCACTCTCTGAAAACAGAAATGAAT 2026  
|||  
601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620  
|||  
2027 CATTCCATCACAATGTGCGAGGAAACAGATTGCGAGTCTTTTAGTAAATGCTTCTCTCTGC 2086  
|||  
621 HisSerIleThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640  
|||  
2087 ACATTTCCACGACGAAATCTAAGATGAATCGAGTATTTCTAGTTATCCACGACATTAAC 2146  
|||  
641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerTyrProAspIleAsn 660  
|||  
2147 TTCAATTCGATTTGAGGACGTTTCATCAAGAAACCCGAGAAACTTAAACGCTCTTC 2206  
|||  
661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680  
|||  
2207 TGTCTTTTGAAGAGTCAACAGAAACCTTACTGGGTGTGTGACATTTACAAGACAG 2266  
|||  
681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700  
|||  
2267 AGCTTTGAAGATTTTCCAGAAATCGGAAGATGTGAAAAACCTTGCACGATTCGATGTC 2326  
|||  
701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720  
|||  
2327 ACTTAGCAAGGTACCATAGAAATAATGCGCAAGGATGTCTACAGTGTGATTTTGCAAT 2386  
|||  
721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740  
|||  
2387 CGTTTGTGAGAGTGTGTAAACAGTGCAGGACTTGTGCAAGAGAAATCCCGCTTTTA 2446  
|||  
741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760  
|||  
2447 ATCAATCCTCAGTTGATTTTTCACGCTCTTTCACGAGGTGTGCGATCACAATGAATGT 2506  
|||  
761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
|||  
2507 CTAATTATACAGGTACTGACGAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566  
|||  
781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800  
|||  
2567 TGTCTCCGAGACCCACGAGTGGAGTGAAGGAGACGACTGCGAGCGCGCTGCACCTGAG 2626  
|||  
801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThrGlu 820  
|||  
2627 ATCGTTGCCATCGATGCTCTTTCACCTTCAGACGCTTACCTCGATCAGTTTGTGCTGAGAAA 2686  
|||  
821 IleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGluLys 840  
|||  
2687 ATGAGACGCGAGCTGAACAAGGCTTACTGTGATTTCTCCGCTCGTGGATTTCTTCAGAG 2746  
|||  
841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860  
|||  
2747 AATCTTTCTGCGTGGCCACAGAGAACTGGGCTGTGCTGCTTGGCTTGGGGGTGATGCCAGG 2806  
|||  
861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880  
|||  
2807 TTAAGACCTTATATACAGATATTGGCAGCTGCTGCAGCTGACGAGATGTGGTTTATTTC 2866  
|||  
881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyrPhe 900  
|||  
2867 ACCTTTGGGAGCTCAGAAATTTGATGAGACATTTTACAGCATGCATTTCTTCTTACTGAA 2926  
|||  
901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920  
|||  
2927 AGGAACTCAGTGTGGAGATGTATAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2986  
|||  
921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940  
|||  
2987 AGAACTGTTCACCCCTGGACACAGACATCAAGCTTTTATCCATTCATATACCATGCTGTC 3046  
|||

Db 941 ArgAsnCysSerThrProGlyProAspIleLeuTyrProPheIleTyrHisAlaVal 960  
 QY 3047 GAGTCTGTGCAGAGACCGCTGACCATTCAGGGCAAGGACGAGCC 3094  
 Db 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
 RESULT 5  
 AAE25629  
 ID AAE25629 standard; protein; 977 AA.  
 XX AC AAE25629;  
 XX DT 04-NOV-2002 (first entry)  
 XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
 XX KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytosstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisenase therapy.  
 XX OS Bos taurus.  
 XX PN US6395543-B1.  
 XX PD 28-MAY-2002.  
 XX PF 23-FEB-2000; 2000US-00511507.  
 XX PR 01-MAY-1998; 98US-0083768P.  
 XX PR 30-APR-1999; 99US-00302812.  
 XX PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-535641/57.  
 XX DR N-PSDB; AAD42081.  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX Claim 3; Col 47-45; 77pp; English.  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG  
 XX Sequence 977 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.31% Indels: 1  
 DB: 5 Gaps: 1  
 US-09-302-812-3 (1-4069) x AAE25629 (1-977)

QY 167 ATGAATCGCGCCCGCTGTGAACCTGTGACCAACGACCCCGTGGCGCGCTACA 226  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTtpAspAlaAla 20  
 QY 227 ACTTCG---CCGCTGCTTCGACGCGCGAGCTTTCGAGCAGCGAGCGGCTCTC 283  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 284 GACCCAGGAGCGCTCAGTCGCTCAGTTCAGGTCCACCGCTCTGCCAGCTCGGTCCA 343  
 Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 344 GCGCAGCGCGGACAGCAGCAGCGCGCCACCTCGCTTGTGTTTCAACAAAAGACTATT 403  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrile 80  
 QY 404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463  
 Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
 QY 464 GAAACACACATACAGAAATAGATCCATGATGATGTTCTGTACAAAAGATAACTTTTAC 523  
 Db 101 GluAsnAsnThrArgGluSerMetMetSerValGlnLysAspAsnPheLys 120  
 QY 524 CAACATAATGTAGAAAATTTAGTAAATGTTTCTCAGCTTAAGTCTTGATTAAGTCACTACT 583  
 Db 121 GlnHisMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 584 GAAAAAGTACACAGTATTGAAACAGCATCAGACTGACGCAATGTGTAAGTGGCAAAAT 643  
 Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 644 GAAGGGAACACACGCGACGCTTTGGAAGTGAACCTCAACACAGTAACCTCGTGTACCA 703  
 Db 161 GluGlyProHisSerGluLeuGluSerGluProProAlaValThrLeuValPro 180  
 QY 704 GAGCAGTTTGTAGTAAATGCTAACATTGATCGGTCAACCTCAAAATGTATGATCAGTACACA 763  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 764 GATAGTGAAGAGATAGAGACAACTCAACAGTTTCTCAGACTGTAAAGCTTCAATGCA 823  
 Db 201 AsnSerGluGluSerArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 824 AAGCAGACTACGGAAGATGAACACGCGCAGAGAAAGCCAAAGCCACGAGAGTGCAGCA 883  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 884 TCTTCCCATCTCGGGAAGACTGTGCAAGTGTGTCAGCAAGATGAGATAGACGTGTGCCA 943  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 944 AAGACTCCATTGTTCAGATGTTGGCTCTCAGGATGTTGGTACTGGGTCAAAAATGACAAC 1003  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 1004 AAATTTGATTAGCAAGAAAGTTGCTTAGGAATTTCTCTCCATTTTGAAGAGAAAGTGA 1063  
 Db 281 ArgLeuAsnArgGlnLysSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 1064 CCCGAATCACCGATGATGTGATTAATTTCTAAATAATGTTTCAAGACTCAGACAGCAGT 1123  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 1124 GAGGAGACAGTCCAGGTTTTCATGACCAAGAGATGGTAGTTCTCCCAAAACAGCAAAAT 1183  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
 QY 1184 AAACCTTCAAGGTTCCAGCAAGACGCTGACATTGAATTTTAGGAAACCGTACTTACT 1243  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360

1244 AAGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGGAGAGAGTGCACCTGGAATG 1303  
 361 LysGlyGlyGluLeuArgLeuHisPheGlnPheGluGlyGlySerArgAlaGlyMet 380  
 1304 AATGATTAAATCTAAATCTCTGGAATATATTTCTAGCTGATCTAGATGACAGAAAT 1363  
 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 1364 TCTAACCAATCGAAAGAGGATCTTAAATACACAGATCATTTGTAGTACAGTGCACAAA 1423  
 401 SerLysGlnHisGlyArgLysAspSerLysLeuLeuHisPheMetArgValProLys 420  
 1424 GCAGAGCAGACAGAAAGAACAGTGGGAAACCAACATCAAAAGACAGAAAGAGATC 1483  
 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 1484 CTTAAATACCTTCCACT 1543  
 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 460  
 1544 GAGATGAGAGAAATGCTCGGTGGGATCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1603  
 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
 1604 CACACAGTAATCTCGGTAGATCTTTTTCGAGCAGGAGAGTCTCTAAACCTTTTCCA 1663  
 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
 1664 ACATTTATTAAGATTGTGGATATCAACAGCATGTTAAATGCTTTTTCAGAACAAAT 1723  
 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 1724 TTGTACCACTGGAAGATGAGATGGTGAGCAACTCTCGGAGCGCGGTGGAGCTCAAT 1783  
 521 LeuTyrProValGluAspGluAsnGlyLysArgAlaAlaGlySerArgTrpGluLeuIle 540  
 1784 CAGACTGCATCTCTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTTGAAA 1843  
 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
 1844 TACATGTGCAATCTTACAAATGGGACTTTTACAGCTTTGATCGATTTCTGGGATAG 1903  
 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
 1904 GTACTTGAAGACAGAGCTCAACATTTATATCAGTCCATCTTGTCTGATATGGTGAAA 1963  
 581 ValLeuGluGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
 1964 ATTGCACTGTCTGCAAAATATTGTCACCGACCAATACACTCTCTGAAACAGAGATG 2023  
 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
 2024 AATCATTTCCATCAATGTCGAGGACAGATGCGAGTCTTTTGTAGTAAATGCTTTCTTC 2083  
 621 AsnHisSerIleThrMetSerGlnGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
 2084 TGACATTTCCACAGCAATGCTAAGATGAAATCGGAGTATTTCTAGTACCCAGACATT 2143  
 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 2144 AACTTCAATCGATTGTTGAGGAGCTTTCATCAAGGAAACCGGAGAACTTAAACCTC 2203  
 661 AsnPheAsnArgLeuPheGluArgLysSerSerArgLysProGluLysLeuLysThrLeu 680  
 2204 TTTGCTACTTTTGAAGAGTTCACAGAGAAACCTACTGCTGGTGGTGGACATTTTACAAGA 2263  
 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 2264 CAGAGTCTTGAAGATTTTCCAGAAATGGAAGATGTGAAACCTTTTACACAGATTGCAT 2323  
 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
 2324 GTCACTTACGAGGATACCATAGAGAAATGCGCAGGCATGCTACAGGTGATTTTGA 2383

721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 2384 AATCGTTTGTGGAGTGGTGTAACTACCACTGACAGACTTGTGCAAGAGAAATCCGCTTT 2443  
 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
 2444 TTAATCAATCTGAGTTGATTATTTCACGCTCTTCACTGAGGTGCTGATCAATGAA 2503  
 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 2504 TCTCTAAATTTATCACAGTACTGACGAGTACAGTGAATACACAGGCTATCTGAGACATAT 2563  
 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 2564 CTTGTCTCCGAGGACCAAGATGGAGTGAAGGAGCACTGCGAGCGGCTGCATC 2623  
 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820  
 2624 GAGATCGTTGCAATCGATCTCTTCACTTCAAGCTTCTGAGCTACTCGATCAGTTTGCCTGAG 2683  
 821 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 840  
 2684 AAAATGAGCGCGCTGCAAGGCTTACTGTGATTTCTCGTCTCTGAGTTTCTTCA 2743  
 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 2744 GAGAATCTTCTGCTGCTGCGCCACAGAAACTGCGGCTGTGGTCTTGGGGTGTATGCC 2803  
 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 2804 AGGTTAAAGCTTATACAGATATTGGCAGCTGCTGCGAGCTTACAGCATTCACATTTTCTTACT 2863  
 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 2864 TTTCACTTTGGGACTCAGAAATGATGAGAGCATTTACAGCATTCACATTTTCTTACT 2923  
 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 2924 GAAAGGAACTCACTCTTGGAGATGTATAGCTTGTCTACGATACATCAATAGAGAA 2983  
 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 2984 TGCAAACTGTTTCCACCTTGGACAGACATCAAGCTTTATCCATTCATATACCATGCT 3043  
 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 3044 GTCGAGTCTCTGTCAGACACCGCTGACCATTCAGGCAAGAGACAGGG 3091  
 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
 RESULT 6  
 AAU76020  
 ID AAU76020 standard; protein; 977 AA.  
 XX  
 AC AAU76020;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Cow; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 OS Bos taurus.  
 XX  
 XX Key Location/Qualifiers  
 FT 601..617  
 FT Region /note= "Represents PARG oligopeptide #1"

FT Region 761..770  
 /note= "Represents PARG oligopeptide #2"  
 FT Region 771..801  
 /note= "Represents PARG oligopeptide #3"  
 FT Region 849..880  
 /note= "Represents PARG oligopeptide #4"  
 XX US6337202-B1.  
 FN 08-JAN-2002.  
 PD 23-FEB-2000; 2000US-00511477.  
 PP 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-163240/21.  
 DR N-PSDB; ABK14931.  
 XX Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX Claim 2; Col 47-52; 81pp; English.  
 XX The present invention relates to a new poly (ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the bovine PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX SQ Sequence 977 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.31% Indels: 1  
 DB: 5 Gaps: 1

US-09-302-812-3 (1-4069) x AAU76020 (1-977)

QY 167 ATGAATGGGGCCCGCTGTGACACCTGCACCAAGACCGCTGGGGCCGCTACA 226  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 QY 227 ACTTCG---CCGCTGCTTCGGACGCCCGAGCTTCCGAGCAGGCGGCGCTC 283  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 284 GACCCCAAGGACCTCACGTGAGTTCAGGTCCACCGCTCTCGCAGCGCTGCTCCA 343  
 Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 344 GGCAGCGCGGACAGCACAGAGCGAGCGGCACCTCGCTGTTTCAAAAGACTATT 403  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80

QY 404 ACCAGTTCGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAA 463  
 Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
 QY 464 GAAACAAACATACAGAAATAGATCCATGATAGTTCCTGTACAAAAAGATTAACCTTTAC 523  
 Db 101 GluAsnAsnAsnThrArgGluSerMetMetSerValGlnLysAspAsnPheTrp 120  
 QY 524 CAACATAATGTAGAAAAATAGTAAATGTTCTCAGCTAAGTCTTGTAAAGTCACTCACT 583  
 Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 584 GAAAAAGTACACAGTATTGTAACCCAGCATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643  
 Db 141 GluLysGlyThrGlnTyLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 644 GAAGGAAACACACGAGCAGCTTTGGAAAGTGAACCTCAACAGTAACTCTGTACCA 703  
 Db 161 GluGlyProHisSerGluArgLeuGluSerGluProProAlaValThrLeuValPro 180  
 QY 704 GAGCAGTTTACTAATGCTTAACATTCGTCACCTCAAAATGATGATCAGACTGACACA 763  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 764 GATAGTGAAGAGAATAGAGACAATCAACAGTTCCTCACAACCTGTAAAGCTTCAAAATGCA 823  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 824 AAGCAGACTACGGAAGATGAACACCCAGAGAGCAAAAGCCACAGNAGTCGACGACAG 883  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 884 TCTTCCCATCTGGGGAAGACTGTGCAAGTTCAGCAAGATGAGATAGACTGCTGGTGCCA 943  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
 QY 944 AAGAGTCCATGTCTCAGATGTTGGCTCTCAGAGATGTTGGTACTGGGTCAAAAATGACAAC 1003  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 1004 AAATTGATTAGCAAGAAAGTTGCTCAGAAATTCCTCCATTTGAGAGAAAGATGAA 1063  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 1064 CCCGAATCACCGATCGATGTGATAATCTCAAAATAGTTGTCAAGACTCAGAAGCAGAT 1123  
 Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
 QY 1124 GAGGAGACAAGTCCAGGTTTGTATCAACAAGAAAGATGGTAGTTCCTCCCAACAGCAAT 1183  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
 QY 1184 AAACCTTCAGGTTCCAGCAGACGCTGCATGTAATTTAGGAAACGGTACTCTACT 1243  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1244 AAGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGGAGAGAGTCCGCTGAAATG 1303  
 Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1304 AATGATTAAATGCTAAACTACCTCGAAATATTTCTAGCCTGAATGTAGAAATGCAGAAAT 1363  
 Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1364 TCTAGCAACATCGAAAAAGAAAGATTTAAATTCAGATTCATTTGATGAGTGCCTCCAAA 1423  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1424 CAGAGAGGACAGAAAGAAAGAACAGTGGGAAACCAACATCAAGAACAGAAAGAGATC 1483  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1484 CCTAAATAGTTCACCTCACCTTCTTCAGATAAAGAGTGGCTTGGAACTCCATTGAG 1543



```

Db      441  ProLysTyrIleProProHisLeuSerProAspLysTyrLeuGlyThrProIleGlu 460
QY      1544  GAGATGAGAAGATCGCTCGGTGGGATCGCGCTCCTCTCTTGAGACATCTGCCAAT 1603
Db      461  GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480
QY      1604  CACACAGTAACATTCGGGTAGATCTTTTCGAGCAGGAGAGAGTCTCTAAACCTTTCCA 1663
Db      481  HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500
QY      1664  ACACATTAATAAGATTGTGGGATAACAAGCATGTTAAATGCTGTGTTTCAGAACAAAT 1723
Db      501  ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520
QY      1724  TTCTACCCAGTGAAGATGAGATGGTGGCGCAACTGCGGGGACCGGTGGGAGCTCAT 1783
Db      521  LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540
QY      1784  CAGACTGCATCTCTCAACAAATTTACGACGCCCAAACTTGAGGATGCTATTTCTGAA 1843
Db      541  GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560
QY      1844  TACAATGTGCATATCTTAAGAATGGGACTTTACAGCTTTGATCGATTTCTCGGATAAG 1903
Db      561  TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580
QY      1904  GTACTTTGAAGACAGAGAGCTCAACATTTATATCAGTCCATCTTGCTGATATGGTGAA 1963
Db      581  ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600
QY      1964  ATTGCACTCTGTGCAAAATATTTCACCCGCCAATACCATCTCTCGAAACAGAGATG 2023
Db      601  IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620
QY      2024  AATCATTTCCATCAATGTCGAGAACAGATGTCAGTCTCTTTAGCTAATGCTTTCTTC 2083
Db      621  AsnHisSerIleThrMetSerGlnGlnGluIleAlaSerLeuLeuAlaAsnAlaPhePhe 640
QY      2084  TGCACATTTCCACGACGAAATGCTTAAGATGAATTCGGAGTATTTCTAGTTACCCAGCAT 2143
Db      641  CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660
QY      2144  AACTTCATCGATTTGTCGGGACCTTCATCAGGAAACCGGAGAACTTAAACGCTC 2203
Db      661  AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680
QY      2204  TTCTGCTACTTTAGAAGAGTCACAGAGAAACCTACTGCGTTGGTGACATTTACAGA 2263
Db      681  PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700
QY      2264  CAGAGTCTTGAAGATTTCCAGATGGGAAAGATGTGAACCCCTTGACAGATTCAT 2323
Db      701  GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720
QY      2324  GTCACTTACGAGGTACCATAGAGAAATGCGCAAGCATGCTACAGGTGGATTTTGCA 2383
Db      721  ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740
QY      2384  AATCGTTTGTGGAGGTGGTGAACCAAGTCAGAGCTGTGCAAGAGAAATCGCTTT 2443
Db      741  AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760
QY      2444  TTAATCAATCTGAGTTGAATTTTACGGCTCTTCACTGAGGTGCTGATCAATGAA 2503
Db      761  LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780
QY      2504  TGTCTAATTTACAGGTACTGAGCAGTACAGTGAATACAGAGGTATGCTGAGACATAT 2563
Db      781  CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800
QY      2564  CTTTGGTCCCGGAGCCAGAGATGGGAGTGAAGGAGCAGCTGCGAGCGCGCTGCACT 2623

```

```

Db      801  ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgArgThrThr 820
QY      2624  GAGATCGTGGCATCGATCTCTTACATTCACAGCTACCTCGATCAGTTGTGCTGAG 2683
Db      821  GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840
QY      2684  AAAATGAGACGGAGCTGAACAGGCTTACTGTGGATTTCTCCGTCTCGAGTTTCTTCA 2743
Db      841  LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860
QY      2744  GAGAACTTTTCTGAGTGGCCACAGAAACTGGGCTGTGGTCCCTTGGGGGTGATGCC 2803
Db      861  GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880
QY      2804  AGGTTAAAGCCCTTAATACAGATATTGGCAGCTGCTGCAGCTGAGCAGATGGGTATT 2863
Db      881  ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900
QY      2864  TTCACCTTTGGGACTCAGAAATTTGATGAGAGACATTTACAGCATGCACATTTCTTACT 2923
Db      901  PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920
QY      2924  GAAAGAAACTCACTGTTGAGATGTGTATAGCTGTTGCTACGATACATACATGAAGA 2983
Db      921  GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940
QY      2984  TGCAGAAACTGTTCCACCCTCGACAGACCGCTGACCATTCAGGCAAGAGCAGGG 3043
Db      941  CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960
QY      3044  GTCGAGTCTCTGTCAGAGACCGCTGACCATTCAGGCAAGAGCAGGG 3091
Db      961  ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 7
AAU75799
ID  AAU75799 standard; protein; 977 AA.
XX
AC  AAU75799;
DT
DE  08-MAY-2002 (first entry)
XX
XX  Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.
XX
XX  Cow; poly (ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;
XX  adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;
XX  inherited genetic disease; myocardial infarction; vascular stroke; aging;
XX  neurodegeneration; Huntington's disease; Parkinson's disease;
XX  Alzheimer's disease; neurotoxicity.
XX
OS  Bos taurus.
XX
FH  Key
FT  Region 601..617
FT  /note= "Represents PARG oligopeptide #1"
FT  Region 761..770
FT  /note= "Represents PARG oligopeptide #2"
FT  Region 771..801
FT  /note= "Represents PARG oligopeptide #3"
FT  Region 849..880
FT  /note= "Represents PARG oligopeptide #4"
XX
XX  US6333148-B1.
XX
XX  25-DEC-2001.
XX
XX  30-APR-1999; 99US-00302812.
XX
XX  01-MAY-1998; 98US-0083768P.
XX  (KENT ) UNIV KENTUCKY RES FOUND.
XX  Jacobson MK, Jacobson EL, Ame J, Lin W;
PI

```



XX WPI; 2002-153820/20.  
 DR N-PSDB; ABK14493.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 XX Claim 3; Col 45-52; 80pp; English.  
 XX  
 CC The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage,  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the bovine PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX  
 XX Sequence 977 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.31% Indels: 1  
 Gaps: 1

US-09-302-812-3 (1-4069) x AAU75799 (1-977)

QY 167 ATGAATGGGGCCCGGGTGTGAACCCCTGCACAAAGCGACCCGTCGGGCGCGCTACA 226  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 QY 227 ACTTCG---CCGGCTGCTTCGACGCGCGGACCTTCCGAGCAGCAGGCGCGCTCCTC 283  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 284 GACCCACAGGAGCTCACCTGAGTTCAGGTCACACCCCTCCGACGCTCCGCTCCA 343  
 Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 QY 344 GGGCAGGCGGAGCAGCAGCAGGCGCGCCACCTCGCTGTTTTCMAACAAAGACTATT 403  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 QY 404 ACCAGTTGATGACACTAAAGGAATCAAGACGAGCGGAATCAGAAAGTTTGGATAGTAAA 463  
 Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerLeuSerLeuHisSerLys 100  
 QY 464 GAAACACAAATCAAGAATAGAAATCCATGATGAGTTCGTACAAAAGATAACTTTTAC 523  
 Db 101 GluAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
 QY 524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGCTTGTATAGTCACTCACT 593  
 Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 584 GAAAAAGTACAGATATTGTAACCGACATCAGATCGCAGCAATGTAGTGGCAAAAT 643  
 Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 644 GAGGGGAAACACACGAGCAGCTTTTGGAAAGTGAACTCAAAACAGTAAACCTTGGTACCA 703  
 Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
 QY 704 GAGCAGTTTGTAGTAACTATGATTCGCTGCTCACCTCAAAATGATGATCAGTGACACA 763

Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 764 GATAGTGAAGAGATAGAGCAATCAACAGTTTCTCAACTGTAAAGCTTGCATATGCA 823  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 824 AAGCAGACTAGGAGATGAACACAGCCAGAGACCCAAAGCCACCCAGAGTGCAGCAAG 883  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 884 TCTTGCCATCTCGGGAAGACTGTGCAAGTTGTGAGCAAGATGAGATAGATGCTGCTGCA 943  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
 QY 944 AAGAGTCCATTTGTCAGATGTTGGTCTGAGGATGTTGGTCTGAGGATGTTGGTCTGAGG 1003  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 1004 AAATTGATTAGACAAGAAAGTTGCTAGGAATTTCTCCATTTGAGAGAAAGTCAA 1063  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
 QY 1064 CCCGAATCACCGATGATGCTGATAATTCTATAAATAGTTCTCAAGATCTCAGAGCAGAT 1123  
 Db 301 ProGluSerProMetAspValAspAsnSerLysSerLysSerCysGlnAspSerGluAlaAsp 320  
 QY 1124 GAGGAGACAGTCCAGGTTTGTATGAACAAGAGATGTTGTTCTCTCCCAACAGCAAT 1183  
 Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
 QY 1184 AAACCTTCAAGTTCCAGCAGAGACCTGACATTTAGTAAATTTAGGAAACGGTACTACT 1243  
 Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
 QY 1244 AAGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGGAGAGTGCACCTGGAATG 1303  
 Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
 QY 1304 AATGATTAAATGCTAAACCTACCTGGAATATTCTACCTGAAATGTAGATGCAAT 1363  
 Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
 QY 1364 TCTAGCAGCACTGNAAGAAAGATTTAAATCACAGATCATTTGATGAGATGCCCCAAA 1423  
 Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
 QY 1424 GCAGAGGACAGAGAAAGAACAGTGGGAAACCAAAACATCAAGAACAGAAAGAGATC 1483  
 Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
 QY 1484 CTTAATAGCTTCCACCTCACCTTCTCCAGATAGAGTGGCTTGGAACTCCCATTTGAG 1543  
 Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 460  
 QY 1544 GAGATGAGAGAAATCCTCGGTGGGATCCGGCTCGCTCTCTTGAGACATCTGCAAT 1603  
 Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
 QY 1604 CACAGATAACTATTTCGGGTAGATCTTTTGCAGCAGGAGAGAGTTCCTAAACCTTTTCA 1663  
 Db 481 HisThrValThrIleArgValAspLeuLeuArgLysGluValProLysProPhePro 500  
 QY 1664 ACACATATAAGATTGTGGGATTAACAGCATGTTAAATGCTTGTTCAGACAAAAAT 1723  
 Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
 QY 1724 TTGTACCCAGTGGAGATCAGATGTGAGCGAACTCGCGGAGCGCGGTGGGAGTCAAT 1783  
 Db 521 LeuTyrProValGluAspGluAsnGlyLysGluAlaGlySerArgTrpGluLeuIle 540  
 QY 1784 CAGACTGCACCTTCTCAACAAATTTTACAGCAGCCCCAAACTTGAAGATGCTATTCTGAAA 1843

541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
1844 TACAATGTGGCATTTCTAAGAAATGGGACTTTACAGCTTTGATCCGATTTCTGGGATAAG 1903  
561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuLeuAspPheTrpAspLys 580  
1904 GTACTTGAAGACAGAGAGCTCAACATTTATATCAGTCCCATCTTGGCTGATATGGTGAAA 1963  
581 ValLeuGluGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
1964 ATTGCACCTCTGTCTGCCAAATATTTCACCCAGCAGCAATACCATCTCTGAAACAGAGATG 2023  
601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
2024 AATCAITTCATCAATGTGCGAGAACAGATGCCAGCTCTTTAGCTAATGCTTCTTC 2083  
621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
2084 TGCACATTTCCACGACCAATGCTAAGATGAATCGGAGTATTCTAGTACCAGACATTT 2143  
641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
2144 AACTTCAATCGATTGTTGAGGGACGTTTCATCAAGGAAACCGGAGAACTTAAACGCTC 2203  
661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
2204 TTCTGCTACTTTAGAGAGTCACAGAGAAACCTACTGCTGGTGGTGCACATTTACAGA 2263  
681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
2264 CAGAGCTTGAAGATTTTCCAGAAATGGAAGATGTGAAGAAACCTTGACACGATTCGAT 2323  
701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
2324 GTCACATTACGAAGGTACCATAGAAGAAATGCCAAGCATGCTACAGTGAGATTTGCA 2383  
721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
2384 AATCGTTTGTGGAGGTGGTGTAAACAGTCAGGACTGTGCAAGAGAAATCGCTTT 2443  
741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
2444 TTAATCAATCTGAGTGCATTTTACGGCTCTTCACTGAGGTGCTCGATCACAATGAA 2503  
761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
2504 TGCTTAATATTACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563  
781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
2564 CGTTGGTCCGAGCCAGCAAGATGGAGTGAAGAGGAGCTGCGAGCGGCGCTGCACT 2623  
801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820  
2624 GAGATCGTTGCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTGAG 2683  
821 GluIleValAlaIleAspAlaLeuHisPheAspArgTyrLeuAspGlnPheValProGlu 840  
2684 AAAATGAGACGGAGCTGAACAGGCTTACTGTGATTTCTCGCTCGCTGAGTCTTCA 2743  
841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
2744 GAGATCTTTGCTGCTGCGCCAGCAAACTGGGGCTGTGGCTTTTGGGGGTGATGCC 2803  
861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyAspAla 880  
2804 AGGTTAAAAGCCTTAATACAGATATTGCGAGCTGCTGCGAGTGCAGAGATGTGGTTAT 2863  
881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
2864 TTCACCTTTGGGACCTCAGAAATTTGATGAGACATTTACAGCATGCACATTTTCTTACT 2923  
901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920

2924 GAAGAGAAACTCACTTTCGAGATGTGTATAGCTTTCTAGCATACTACATGAGAA 2983  
921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuArgTyrTyrAsnGluGlu 940  
2984 TGCAGAAACTGTTCCACCCCTGGAGCCAGACATCAAGCTTTATCCATTCATATACCATGCT 3043  
941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
3044 GTGAGTCTCTGTGCAGAGACCGCTGACCATTCAGGCGCAAGGACAGG 3091  
961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
RESULT 8  
ABG72278  
ID ABG72278 standard; protein; 977 AA.  
XX ABG72278;  
AC ABG72278;  
XX 13-MAR-2003 (first entry)  
XX Bovine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
DE Bovine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
KW cellular response; DNA damage; neoplastic disorder inducing agent;  
KW oxidative stress; neoplastic disorder; myocardial infarction;  
KW Parkinson's disease; neurodegenerative disease; Alzheimer's disease;  
KW vascular stroke; Huntington's disease; inborn genetic error;  
KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
KW neuroprotective; nontropic; antiparkinsonian; cardiac; vasotropic;  
KW anticonvulsant; cerebroprotective; enzyme.  
XX Bos taurus.  
XX US2002132328-A1.  
XX 19-SEP-2002.  
XX 09-OCT-2001; 2001US-00973451.  
XX 01-MAY-1998; 98US-0083768P.  
XX 30-APR-1999; 99US-00302812.  
XX (JACO/) JACOBSON M K.  
XX (JACO/) JACOBSON E L.  
XX (AMEJ/) AME J.  
XX (LINW/) LIN W.  
XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
XX WPI; 2003-155895/15.  
XX N-PSDB; ABX14477.  
XX New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose)  
XX glycohydrolase activity, for preventing, treating, or ameliorating a  
XX disease condition, e.g. neoplastic disorder, myocardial infarction or  
XX vascular stroke.  
XX Claim 28; Fig 16; 86pp; English.  
XX The present invention relates to the isolation of poly (ADP-ribose)  
XX glycohydrolase (PARG) from several species, and the polynucleotide  
XX sequences encoding them. Methods for inhibiting PARG expression or  
XX overexpressing PARG are also disclosed. PARG is involved in the cellular  
XX response to DNA damage, and is associated with the body's response to  
XX neoplastic disorder inducing agents and oxidative stress. The  
XX polynucleotide sequences encoding PARG and PARG modulators are useful for  
XX preventing, treating, or ameliorating diseases such as neoplastic  
XX disorders, myocardial infarction, vascular stroke, neurodegenerative  
XX disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
XX disease), inborn genetic errors, reperfusion following ischaemia, aging,  
XX and neurotoxicity. The polynucleotide sequences are also useful in gene  
XX therapy. The methods are useful for identifying an agent that modulates

CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme

XX  
 SQ Sequence 977 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.31% Indels: 1  
 DB: 6 Gaps: 1

US-09-302-812-3 (1-4069) x ABG72278 (1-977)

QY	167	ATGATGCGGCGCGCGCTGTGAACCCCTGCACAAAGCGACCGCTGGGCGCGCTACA	226
DB	1	MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTyrAspAlaAla	20
QY	227	ACTTCG---CCGCGCTCTCGACGCCCGAGCTTTCGAGCAGGCGCGCTCTC	283
DB	21	ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValIleu	40
QY	284	GACCCCAAGGAGCTCAGCTGAGTTCAGGTCGCCAGCTCTCCAGCGCTGCCA	343
DB	41	AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu	60
QY	344	GGCGAGGCGGACAGCAGAGCGCGCGCTCTGCTGCTTTTCAAAACAAAGACTATT	403
DB	61	GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle	80
QY	404	ACCAGTTGGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTGGATGATAA	463
DB	81	ThrSerTipMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys	100
QY	464	GAATAACAATCAAGATAGNATCCATGATGCTTGTACAAAAGATACTTTTAC	523
DB	101	GluAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr	120
QY	524	CAACATAAGTGAAAGATAGTAATGTTTCTCAGCTAAGTCTTGATAGTCACTCACT	583
DB	121	GlnHisAsnMetGluLysValLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal	140
QY	584	GAATAAGTACACAGTATTGTAACCGACATCACTGCGACGATGTGTAGTGGCAAT	643
DB	141	GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysThrGlnAsn	160
QY	644	GAAGGGAACACACGGGAGCAGCTTTTGGAAAGTGAACTCAACAGTAAACCTGTGACCA	703
DB	161	GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro	180
QY	704	GAGCAGTTTATGTAATGTAACATTGATCGGTCACTCAAAATGATGATCACTGACACA	763
DB	181	GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr	200
QY	764	GATAGTGAAGAGATAGAGACAATCAACAGTTTCTCAACTGTAAAGCTTGCAATGCA	823
DB	201	AsnSerGluGluSerArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla	220
QY	824	AAGCAGACTACGGAGATGACACCGCAGAGAGCCAAAGCCACCAAGATGCGACGAA	883
DB	221	LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys	240
QY	884	TCTTGCATCTCTGGGAGAGCTGTGCAAGTTGTGCAAGATGATGATGATGCGTGC	943
DB	241	AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer	260
QY	944	AAGAGTTCATTTGATGTTGGCTCTGAGATGTTGGTACTGGGTCAAAAATGACAC	1003
DB	261	GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn	280
QY	1004	AAATTGATTAGACAAAGATTGCTAGGAATTTCTCTCCATTTGAGAGAAAGTGAA	1063

DB	281	ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu	300
QY	1064	CCCGAATCACCGATGGATGTGATAATTTCTAAATAATAGTTCTCAAGACTCAGAAGCAGAT	1123
DB	301	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	320
QY	1124	GAGGAGCAAGTCCAGGTTTGTGACAAAGAGATGGTAGTCTCTCCCAACAGCAAT	1183
DB	321	GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn	340
QY	1184	AAACCTTCAAGGTTCCAAAGCAAGAGACCTGACATTTGAATTTAGAAACGGTACTCTACT	1243
DB	341	LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla	360
QY	1244	AAGGCGGTGAGTTAGATTACATTTCCAAATTTGAAGGAGAGAGAGTCCGACTGGAATG	1303
DB	361	LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet	380
QY	1304	AATGATTTAAATGCTTAAACTTACCTGGAATATTTCTAGCCTGAATGTAGATGCAAT	1363
DB	381	AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn	400
QY	1364	TCTAAGCAACATGGAAGAAAGATTTCTAAATCACAGATCATTTGTGATGAGCTGCCAAA	1423
DB	401	SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys	420
QY	1424	GCAGAGGACAGAGAAAGAACAGTGGGAACCAACATCAAGAACAGAAAGAGATC	1483
DB	421	AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle	440
QY	1484	CCTAATAAGCTTCCACCTCACCTTTCTCAGATAAGAGAGTGGCTGGAACTCCATTGAG	1543
DB	441	ProLysTyrIleProProHisLeuSerProAspLysLysTyrLeuGlyThrProIleGlu	460
QY	1544	GAGATGAGAAGATGCTCGGTGGGATCCGCTCTCTGTGAGACCATCTGCGCAAT	1603
DB	461	GluMetArgArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn	480
QY	1604	CACAGTAACATTCGGGTAGATCTTTTCCAGACAGGAGAGAGTCTCTAAACCTTTTCCA	1663
DB	481	HisThrValThrIleArgValAspLeuLeuArgLysGlyGluValProLysProPhePro	500
QY	1664	ACACATTTATAAGATTTGGGATACCAAGCATGTTTAAATGCTTGTTCAGACAAAT	1723
DB	501	ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	520
QY	1724	TGTACCCAGTGGAGATGAGATGAGCGAAGTGGCGGAGCGGCGGTGGAGCTCAT	1783
DB	521	LeuTyrProValGluAspGluAsnGlyGluArgAlaGlySerArgTyrGluLeuIle	540
QY	1784	CACACTGCATCTCTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAAA	1843
DB	541	GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys	560
QY	1844	TACATGTGCAATTTCTAAGAAATGGACTTTTACAGCTTTGATCGATTTCTGGGATAAG	1903
DB	561	TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuIleAspPheTrpAspLys	580
QY	1904	GTACTTGAAGACAGACAGCTCAACATTTATATCAGTCCATCTTGTGCTGATATGGTGAAA	1963
DB	581	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys	600
QY	1964	ATTGCACTCTGTGCGCAATATTTCACCCAGCCCAATACCATCTCTCGAAGACAGATG	2023
DB	601	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	620
QY	2024	AATCATTTCCATCAATGTGCGAGGAACAGATTCCCGAGTCTTTTAGCTAATGCTTCTTC	2083
DB	621	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	640
QY	2084	TGCATTTTCCACGACGCAATGTAGATCAATCGAGATTTCTAGTATCCCAAGACATT	2143

Db 641 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
 QY 2144 AACTTCATCGATTGTTGAGGACGCTTCATCAAGGAACCGGAGAACTTAAACGCTC 2203  
 Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
 QY 2204 TTCTGCTACTTTAGAGAGTACAGAGAAACCTACTCGGTTGGTGACATTTACAGA 2263  
 Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
 QY 2264 CAGAGCTTCAAGATTTCCAGATGGAAGATGGAAGATGGAAGATGGAAGATGGAAGAT 2323  
 Db 701 GluSerLeuGluAspPheProGluTyrPheGluGlyArgSerSerArgLysLeuLysThrLeu 720  
 QY 2324 GTCACTTACGAAGTACCATAGAAATGGCAAGCATGCTACAGGTGGATTTTGA 2383  
 Db 721 ValThrTyrGluGlyThrLeuGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 QY 2384 AATCGTTTGTGGAGTGTGTGTAACAGTGCAGGACTTGTGCAAGAAATCCGCTTT 2443  
 Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 760  
 QY 2444 TTAATCAATCTGAGTTGATTATTTCACGCTCTTCACTGAGGTGCTGATCAATGAA 2503  
 Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspPheAsnGlu 780  
 QY 2504 TGTCTAATTATCACAGTACTGACAGTACAGTGAATACAGAGTATGCTGAGACATAT 2563  
 Db 781 CysLeuIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2564 CGTTGGTCCCGACCCAGAGTGGAGTGAAGAGGACGACTCGAGCGCGCTGCACT 2623  
 Db 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgArgThrThr 820  
 QY 2624 GAGATCGTTGCATGCTGCTCTCACTTCACTGAGCTACCTGATCTGCTGCTGCTGAG 2683  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 QY 2684 AAAATGAGACGCGAGCTCAACAGAGCTTACTGTGATTTCTCCGTCCTGGAGTTCTTCA 2743  
 Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 QY 2744 GAGATCTTTCTGAGTGGCCACAGGAACTGGGGCTGTGTGCTGCTTGGGGGTGATGCC 2803  
 Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 QY 2804 AGGTTAAAGCTTAATACAGATATTGCGAGTGTGCGAGTGTGCGAGTGTGCGAGTGT 2863  
 Db 881 ArgLeuLysAlaLeuIleGlnLeuAlaAlaValAlaGluArgAspValValTyr 900  
 QY 2864 TTCACCTTTGGGACTCAGAATTGTATGAGACATTTACAGATGACATTTTCTTACT 2923  
 Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 QY 2924 GAAAGAACTCCTGTTGGAGATGTATAGCTGTTCAGTATCTACATCAATCAAGAA 2983  
 Db 921 GluArgLysLeuThrValGlyGlyValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 QY 2984 TGCAGAACTGTTCCACCCCTGACAGACATCAAGCTTTATCCATTCATATACCATGCT 3043  
 Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 QY 3044 GTCGAGTCTCTGCGAGAGCCGCTGACCAATTCAGGCGAAAGACAGGG 3091  
 Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976  
 RESULT 9  
 ID AAE25631  
 XX AAE25631 standard; protein; 968 AA.  
 AC AAE25631;  
 XX AAE25631;  
 DT 04-NOV-2002 (first entry)

XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 DE Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
 KW cytosolic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
 XX antisense therapy.  
 OS Mus musculus.  
 XX US6395543-B1.  
 PN 28-MAY-2002.  
 XX 23-FEB-2000; 2000US-00511507.  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42083.  
 DR New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX Claim 3; Col 63-68; 77pp; English.  
 PS The invention relates to an isolated nucleic acid molecule which encodes  
 XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyzes release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG  
 XX Sequence 968 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.57% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: 5 Gaps: 6  
 US-09-302-812-3 (1-4069) x AAE25631 (1-968)  
 QY 167 ATGAATGGGCGCCCGCTGTGAACCTTGCACCAAGCGACCCGCTGGGCGCGCT--- 223  
 Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGGTGCTTCGAGCGCCGCGAGCTTCGAGCAGCAGGCGGCTCCTC 283  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 284 GACCCAGGAGCGCTCAGTTCAGGTCCACCGTCTCGCCAGCTGCGTCCCA 343  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59



Db 772 CysLeuIleThrGlyThrGluInTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
 QY 2564 CGTGTGTCGGAGCCAGAGATGGAGTGAAGGAGCAGCTGCAGCTGCACCT 2623  
 Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluAspThrGlnArgA-CysThr 811  
 QY 2624 GAGATCGTCCGATCGTCTCTTCACTTCAGAGCTTACCTCGATCAGTTGTGCTGAG 2683  
 Db 812 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 831  
 QY 2684 AAATGAGACGGAGCTGAACAGGCTTACTGTGGATTCTCCGTCTGAGATTCTTCA 2743  
 Db 832 LysValArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
 QY 2744 GAGATCTTTCTGAGTGGCCACAGGAACTGGGGCTGTGCTGCTTGGGGGTGATGCC 2803  
 Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
 QY 2804 AGTTAAAGCTTAATACAGATATTCGACGTGCTGAGTGGAGATGTGTTTAT 2863  
 Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891  
 QY 2864 TTCACTTTGGGACTCAGATTCATGAGACATTTACAGCATCTTCTTCTACT 2923  
 Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
 QY 2924 GAAAGGAACTCACTGTGGAGATGTCTATAAGCTGTGTGCTTACCATCTACATGAAGAA 2983  
 Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuA-GTyrTyrAsnGluGlu 931  
 QY 2984 TGCAGAACTGTTCACCCCTGGACACATCAAGCTTTATCCATTCATATACCATGCT 3043  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3044 GTCAGTCTGTGAGACCGCTGACCATTCAGGCGAAGGACAGGAGCC 3094  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968  
 RESULT 10  
 AAU76022  
 ID AAU76022 standard; protein; 968 AA.  
 AC  
 AC AAU76022;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Mus musculus.  
 XX  
 PN US6137202-B1.  
 XX  
 PD 08-JAN-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511477.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W,  
 PI  
 XX WPI; 2002-163240/21.  
 DR  
 DR N-PSDB; ABK14933.  
 XX

PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP-ribose polymer, useful for  
 treating neoplastic and neurological disorders, heart attack and stroke.

Claim 2; Col 63-70; 81pp; English.

XX The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the mouse PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX

Sequence 968 AA;

Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: Gaps: 6

US-09-302-812-3 (1-4069) x AAU76022 (1-968)

QY 167 ATGAATCGCGGCGCGCTGTGAACCTGCACCAAGCGACCGCTGGGCGCGCT--- 223  
 Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLysAla--ArgTrpGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGGCTGCTTCGAGCGCGGAGCTTCCGAGCAGGCGAGCGGCTGCTC 283  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 284 GACCCGAGGACGCTCACGTGCAGTTCAGGCTCCACCGCTCCGAGCTGCTGCCA 343  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
 QY 344 GGGCAGCGCGGACAGACAGAGCGAGCGCGCACCTCGCTTGTTCATCAAAAGACTATT 403  
 Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 404 ACCAGTTGGATGGACACTAAAGGAATCAAGACGCGGAATCAGAAAGTTGGATAGTAAA 463  
 Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 464 GAAACACATACAAAGATAGATCCATGATGATGTTCTGACAAAAGATAAATTTTAC 523  
 Db 97 GluAsnAsnThrThrGlyLeuAspSerMetSerSerValGlnLysAspAsnPheTyr 116  
 QY 524 CAACATAATGTAGAAAAATAGTAAATGTTTCTCAGCTTAAGTCTTGATTAAGTCACTACT 583  
 Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 584 GAAAAAGTACACAGTATTTCAGCAGCATCAGACTCAGCAAGTGTAGTGCAGAAAT 643  
 Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysLysTrpGlnAsn 156  
 QY 644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGATAACCTCGTACCA 703  
 Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 704 GAGCAGTTTGTAGTAAATGCTTACATTCATCGGTGCTCACTCAAAATGATGATCAGTAC 763  
 Db 177 LysGlnLeuSerAsnAlaAsnIleGlnSerProHisThrAspAspHisSerAspThr 196





Db 912 GluArgLysLeuAapValGlyLysValTyrLysLeuLeuArgTyrTyrAsnGluGlu 931  
 QY 2984 TGCAGAACTGTTCCACCCCTGGACACATCATAGCTTTATCATTCATATACCATGCT 3043  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3044 GTCAGTCTGTGACAGACCGCTGACCATTCAGGCGCAAGGACAGGAGC 3094  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 11  
 AAU76013  
 ID AAU76013 standard; protein; 968 AA.  
 XX AAU76013;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX  
 OS Mus musculus.  
 XX  
 FN US6333148-B1.  
 XX  
 PD 25-DEC-2001.  
 XX  
 PF 30-APR-1999; 99US-00302812.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 XX  
 PA (KENT) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-153820/20.  
 DR N-PSDB; ABK14495.  
 XX  
 PT Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 PT useful potentially for treating diseases associated with DNA damage, e.g.  
 PT cancer.  
 XX  
 PS Claim 3; Col 63-68; 80pp; English.  
 XX

The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the mouse PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention

SQ Sequence 968 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: 5 Gaps: 6

US-09-302-812-3 (1-4069) x AAU76013 (1-968)  
 QY 167 ATGAATGCGGCGCGCGCTGTGAACCTTGACCAACAGCGACCGCTGGGCGCGCT--- 223  
 Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGCGCTGCTTCGAGCGCGCGGAGCTTCCGAGCAGCGCAGCGCGCTCCTC 283  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 284 GACCCCAAGGACGCTCACGTGAGTTCAGGTCACCGCTCCCGCTCCGCGCAGCTGCTGCCA 343  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerProAlaCysValSer 59  
 QY 344 GGGCAGCGCGGACGACGACGAGCGCGCCACCTCGCTTCTTTTCAACAAAAGACTATT 403  
 Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 404 ACCAGTTGGATGGACACTAAAGGAATCAACAGCAGCGGAATCAGAAAAGTTTGGATAGTAA 463  
 Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 464 GAAACACACATACAGATAGTATCCATGATGAGTCTGTACAAAAGATAGTACTTTAC 523  
 Db 97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
 QY 524 CAACATAATGTAGAAAAATAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583  
 Db 117 ProHisLysValGlnLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 584 GAAAAAGTACACAGTATTTGAACCGACATCAGACTGCAGCAATGTGTAGTGGCAAAAT 643  
 Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
 QY 644 GAAGGAAACACACGCGAGCGCTTTTGGAAAGTGACCTCAACAGCTAACCCCTGGTACCA 703  
 Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 704 GAGCAGTTTAGTAAATGCTAACATTGATCGGTCACTCAAAATGATGATCACAGTGACACA 763  
 Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 764 GATAGTGAAGAGATAGACAATCAACAGTCTTCTCAAACTGTAAAGCTTGCAAAATGCA 823  
 Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 QY 824 AAGCAGACTACGGAAGTGAACAGCGCAGAGAGCCCAAGCCACAGAGAGTGCAGCAAG 883  
 Db 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233  
 QY 884 TCTTGGCCATCTCGGGAAGACTGTGCAAGTTGTGACAGATGAGATAGACGTGGTGCCA 943  
 Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGlnGluValAspValLeuPro 253  
 QY 944 AAGAGTCCATTGTCAGATGTTGCTCTGAGGATGTTGTTGCTACTGGTCAAAAATGACAAC 1003  
 Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1004 AAATTGATTAGACAAGAAAGTTCCTAGGAAATTCCTCTCATTTGAGAGAGGAAAGTGA 1063  
 Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProGluLysGluSerGlu 293  
 QY 1064 CCCGAATCACCGATGATGAGTAAATTCCTAAATAAGTTGTCAAGACTCAGAGCAGAT 1123  
 Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
 QY 1124 GAGGAGACAAGTCCAGGTTTGTATGAAACAAGAGATGGTGTCTCTCCCAACAGCAAT 1183  
 Db 314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332  
 QY 1184 AAACCTTCAGGTTCCAGCAGAGCGCTGACATGAATTTAGGAAACGGTACTCTACT 1243



Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuLeuArgGlyTyrLeuThr 352  
 QY 1244 AAGGGCGGTAGAGTATCATTTTCAATTTGAAGGAGGAGAGTCCGACTGGAATG 1303  
 Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaGlyThr 371  
 QY 1304 AATGATTTAAATGCTAACTACTCTGGAATATTTCTAGCCTGAATGTAGAATGCAAAAT 1363  
 Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 QY 1364 TCTAAGCAATGAGGAAAAAGGATTTCTAAATTCACAGATCATTTGATGAGACTGCCCAAA 1423  
 Db 392 SerLysGlnHisGlyArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
 QY 1424 CAGAGGACAGAGAAAGAACAGTGGGAACCAACATCAAGAACAGAGAAAGGAGATC 1483  
 Db 412 SerGluAspArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 QY 1484 CTTAATACGTTCCACCTCACCTTTCTCCAGATAAGAGTGGCTTGGAACTCCCATTCAG 1543  
 Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysTyrLeuGlyThrProIleGlu 451  
 QY 1544 GAGATGAGAAGATGCTCGGTGGGATCCGGCTGCTCTTTCAGACCATCTGCCAAT 1603  
 Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
 QY 1604 CACAGTAATCATTTCCGGTAGATCTTTTCGAGCAGGAGAGTTCCTAAACCTTTTCCA 1663  
 Db 472 HisThrValThrValArgValAspLeuLeuArgGluValProLysProPhePro 491  
 QY 1664 ACACATATTAAGATTTGGGATTAACAGCATGTTAAATGCTTGTTCAGAAACAAAT 1723  
 Db 492 ThrHisTyrLysAspLeuTyrAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 QY 1724 TTGTACCCAGTGAAGATGAGATGGTGAGCGAATCTGCGGAGCGCGGTGGAGTCTATT 1783  
 Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrGluLeuIle 531  
 QY 1784 CAGACTGCATCTCAGCAAAATTTACAGACCCCAAACTTCAAGGATGCTATTCTGAAA 1843  
 Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
 QY 1844 TACAATGTGGCATTTCTAAGAAATGGGACTTTACAGCTTTTGATCAATTTCTGGGATAAG 1903  
 Db 552 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuValAspPheTyrAspLys 571  
 QY 1904 GTACTGGAAGACGAGAAGCTCAACATTTATATCATCTGCTCATCTGCTGATATGGTGAAA 1963  
 Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
 QY 1964 ATTGCATCTGTCTGCCAATATTTCCACCCCAATACCACTCTCTGAAACAGAGATG 2023  
 Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
 QY 2024 AATCATTTCCATCACAATGTCGAGGAAACAGATGTCAGTCTTTTAGCTAATGCTTTCTTC 2083  
 Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 QY 2084 TGCACATTTCCAGCAAGAAATGCTAAGATGAATTCGGAGTATTTCTAGTTACCCAGACATT 2143  
 Db 632 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
 QY 2144 AACTTCAATCGATTGTTGAGGACGTTTCATCAAGAAACCGAGAAACTTAAACCGCTC 2203  
 Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
 QY 2204 TTCTGCTACTTTAGAAGAGTCAAGAGAAAAACCTACTGGTGGTGGATTTACAGATTTACAGA 2263  
 Db 672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691  
 QY 2264 CAGAGCTTTGAAGATTTTCCAGAAATGGGAAGATGTGAAAACCCCTTGACACGATTGGAT 2323  
 Db 692 GlnSerLeuGluAspPheProGluTyrProGluArgCysGluLysProLeuThrArgLeuHis 711

QY 2324 GTCACCTACGAGGTACCATAGAGAAATGCGCAAGCATCTACTAGTGGATTTTGA 2383  
 Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
 QY 2384 AATCGTTTGTGGAGGTGGTGAACAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2443  
 Db 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluLysLeuArgPhe 751  
 QY 2444 TTAATCAATCTGAGTTCATTTTACGGCTCTTCACTGAGGTCTGTGATCACAATGAA 2503  
 Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisGlnGlu 771  
 QY 2504 TGTCTAAATATCACAGGTACTGAGCAGTACAGTGAATATCACAGGTCTGTGAGCATAT 2563  
 Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
 QY 2564 GCTTGGTCCCGAGGCCACGAGATGGGAGTGAAGGAGGAGTGCAGCGGGCGTGCAT 2623  
 Db 792 ArgTyrAlaArgSerHisGluAspGlySerGluLysAspAspTyrGlnArgCysThr 811  
 QY 2624 GAGATCGTGGCCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTCGTGCTGAG 2683  
 Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
 QY 2684 AAAATGACGCGAGCTGAAACAGGCTTACTGTGATTTCTCCGCTCTGGAGTTTCTTCA 2743  
 Db 832 LysValAlaArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
 QY 2744 GAGAACTTTCTGAGTGGCCACAGAAACTGGGGCTGTGGTCCCTTGGGGGTGATGCC 2803  
 Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla 871  
 QY 2804 AGTTAAAGCCCTTAATACAGATATGGCAGCTGCTGAGCTGAGCAGATGTGGTTTAT 2863  
 Db 872 ArgLeuLysAlaLeuIleGlnLeuAlaAlaAlaIleValArgAspValValTyr 891  
 QY 2864 TTTCACTTTGGGACTCAGATTTGATGAGACATTTACAGATGACATTTTCTTACT 2923  
 Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
 QY 2924 GAAAGAAACTCAGTGTGGAGATGTATAGCTGTGTCTAGTACTACATGATGAGAA 2983  
 Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
 QY 2984 TGCAGAAACTGTTCACCCCTCGACAGACATCAAGCTTTATCCATTTATATATACCATGCT 3043  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3044 GTCGAGTCTGTGAGAGACCGCTGACCATTCAGGGCAAGACAGGAGACC 3094  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968  
 RESULT 12  
 ABG72280  
 ID ABG72280 standard; protein; 968 AA.  
 XX AC ABG72280;  
 XX DT 13-MAR-2003 (first entry)  
 XX DE Mmurine poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX KW Murine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;  
 KW cytosolic; neuroprotective; nontoxic; antiparkinsonian; cardiac;  
 KW vasotropic; anticonvulsant; cerebroprotective; enzyme.  
 OS Mus musculus.

XX US2002132328-A1.  
 XX 19-SEP-2002.  
 XX 09-OCT-2001; 2001US-00973451.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX 30-APR-1999; 99US-00302812.  
 XX (JACO/) JACOBSON M K.  
 XX (JACO/) JACOBSON E L.  
 XX (AMEJ/) AME J.  
 XX (LINW/) LIN W.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2003-155895/15.  
 XX N-PSDB; ABX14479.  
 XX New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 XX glycohydrolase activity, for preventing, treating, or ameliorating a  
 XX disease condition, e.g. neoplastic disorder, myocardial infarction or  
 XX vascular stroke.  
 XX Claim 28; Fig 16; 86pp; English.  
 XX The present invention relates to the isolation of poly(ADP-ribose)  
 XX glycohydrolase (PARG) from several species, and the polynucleotide  
 XX sequences encoding them. Methods for inhibiting PARG expression or  
 XX overexpressing PARG are also disclosed. PARG is involved in the cellular  
 XX response to DNA damage, and is associated with the body's response to  
 XX neoplastic disorder inducing agents and oxidative stress. The  
 XX polynucleotide sequences encoding PARG and PARG modulators are useful for  
 XX preventing, treating, or ameliorating diseases such as neoplastic  
 XX disorders, myocardial infarction, vascular stroke, neurodegenerative  
 XX disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 XX disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 XX and neurotoxicity. The polynucleotide sequences are also useful in gene  
 XX therapy. The methods are useful for identifying an agent that modulates  
 XX PARG activity, identifying a mutant PARG allele in an individual, or  
 XX screening candidate molecules for PARG modulating activity. The present  
 XX sequence represents murine PARG enzyme  
 XX Sequence 968 AA;  
 Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: 6 Gaps: 6  
 US-09-302-812-3 (1-4069) x ABG72280 (1-968)  
 QY 167 ATCAATGCGGGCCCGCTGTGACCTGCACCAAGCGACCGCTGGGCGCGCT--- 223  
 Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGCGCTGCTTCGAGCGCGCGAGCTTTCCGAGCAGGCGCGCGCTCCTC 283  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39  
 QY 284 GACCCCAAGACGCTACGTCAGTTCAGGTCCTCCACCGCTCCGCGCGCTCGCTCCCA 343  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
 QY 344 GGCAGCGCGGACGACGACGACGCGCGCTCGCTGCTTTTCAACCAAGAGACTATT 403  
 Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 404 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATGATAA 463

80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 464 GAAACCAACATACAAAGATAGATCCATGATGAGTTCGTGTACAAAGATACCTTTTAC 523  
 Db 97 GluAsnAsnAsnThrArgIleAspSerMetSerMetSerValGlnLysAspAsnPheTyr 116  
 QY 524 CAACATAATAGAAAATTAATAGTTCCTCAGCTAAGTCTTGATAAGTCACTCACT 583  
 Db 117 ProHisLysValGluLysLeuAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 584 GAAAAAGTACACAGTATTGTAACCCAGCATCAGACTGCAGCAATGTGTAAAGTGGCAAAAT 643  
 Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
 QY 644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTCGTACCA 703  
 Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 704 GAGCAGTTTAGTAATGCTAACATTGATCGTCACTCAAAATGATGATCAGATGACACACA 763  
 Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 764 GATAGTGAAGACATAGACAAATCAACAGATTCTCACAACCTGTAAAGCTTGCACAAATGCA 823  
 Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 QY 824 AGCAGACTACGAGATGACACAGCCAGAGAGCCAAAGCCACAGAGTGCACAGCAAG 883  
 Db 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233  
 QY 884 TCTTGCCATCTCTGGGAGACAGCTGCAAGTGTCTCAGCAAGATGAGATAGACGTGGTGCCA 943  
 Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
 QY 944 AAAGTCCATTCAGATTGGCTCTGAGAGTGTGGTCTGGGTCAAAAATGACAAAC 1003  
 Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1004 AAATGATTAGCAGAAAGTTCCTAGGAATTCCTCCATTTCAGAGAGAAAGTGAAT 1063  
 Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGlnLysGluSerGlu 293  
 QY 1064 CCCGAATCACCGATGATGATTAATCTAAATAATAGTTGTCAAGACTCAGAGCAGAT 1123  
 Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
 QY 1124 GAGGAGACAAAGTCCAGGTTTTCATGAAACAGAGATGAGTTCCTCCCAACAGCAAT 1183  
 Db 314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332  
 QY 1184 AAACCTTCAAGGTTTCCAGCAGAGACGCTGACATTTAGGAATTTAGGAAACGGTACTTACT 1243  
 Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
 QY 1244 AAGGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGAGAGAGTCCGCTGGAATG 1303  
 Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaAlaGlyThr 371  
 QY 1304 AATGATTAAATGCTAACTACTCTGGAATATTTCTAGCTGAATGATAGATGACAGAAAT 1363  
 Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 QY 1364 TCTAGCAACATGGAATAAGGATTTCTAAATTCACAGATCATTTGATGAGACTGCCCAAA 1423  
 Db 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysLys 411  
 QY 1424 GCAGAGGACAGAAAGAAAGACAGTGGGAAACCAACATCAAGAACAGAAAGAGATC 1483  
 Db 412 SerGluAspArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 QY 1484 CCTAATAGCTTCCACCTCACCTTCTCCAGATAGAAAGTGGCTTGGAACTCCCTGAG 1543











GenCore version 5.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: May 26, 2004, 16:27:48 ; Search time 34.0417 Seconds  
(without alignments)  
12341.705 Million cell updates/sec

Title: US-09-302-812-3  
Perfect score: 7208  
Sequence: 1 ggcgtctgggaagtggagg.....agaaaaaaaaaaaaaaaaaaaa 4069

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 389414 segs, 51625971 residues  
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p.model -DEV=xlpl  
-Q=/cgn2\_1/USPTO\_spool\_p/633148/runat\_26052004\_150054\_5717/app\_query.fasta\_1.12437  
-DB=Issued Patents AA -QFMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148@cgn\_1\_131@runat\_26052004\_150054\_5717 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DETIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5190	72.0	976	4	US-09-302-812-4
2	5190	72.0	976	4	US-09-511-477-4
3	5190	72.0	976	4	US-09-511-507-4
4	4635.5	64.3	977	4	US-09-302-812-2
5	4635.5	64.3	977	4	US-09-511-477-2
6	4635.5	64.3	977	4	US-09-511-507-2
7	4395	61.0	968	4	US-09-302-812-6
8	4395	61.0	968	4	US-09-511-477-6
9	4395	61.0	968	4	US-09-511-507-6
10	1081	15.0	768	4	US-09-302-812-8
11	1081	15.0	768	4	US-09-511-477-8
12	1081	15.0	768	4	US-09-511-507-8

13	463	6.4	726	4	US-09-302-812-10	Sequence 10, Appl
14	463	6.4	726	4	US-09-511-477-10	Sequence 10, Appl
15	463	6.4	726	4	US-09-511-507-10	Sequence 10, Appl
16	167	2.3	31	4	US-09-302-812-11	Sequence 11, Appl
17	167	2.3	31	4	US-09-511-477-11	Sequence 11, Appl
18	167	2.3	31	4	US-09-511-507-11	Sequence 11, Appl
19	154	2.1	1972	4	US-08-875-435B-3	Sequence 3, Appli
20	152	2.1	1231	4	US-09-595-684B-23	Sequence 23, Appl
21	150.5	2.1	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
22	150.5	2.1	1972	4	US-08-875-435B-4	Sequence 4, Appli
23	150	2.1	3969	3	US-08-061-376-5	Sequence 5, Appli
24	149.5	2.1	1187	1	US-08-320-559-28	Sequence 28, Appl
25	149.5	2.1	1187	3	US-08-545-860D-28	Sequence 28, Appl
26	149.5	2.1	1187	5	PCT-US94-04496-28	Sequence 28, Appl
27	149.5	2.1	1210	1	US-08-320-559-26	Sequence 26, Appl
28	149.5	2.1	1210	3	US-08-545-860D-26	Sequence 26, Appl
29	149.5	2.1	1210	5	PCT-US94-04496-26	Sequence 26, Appl
30	149	2.1	26	4	US-09-302-812-20	Sequence 20, Appl
31	149	2.1	26	4	US-09-511-477-20	Sequence 20, Appl
32	149	2.1	26	4	US-09-511-507-20	Sequence 20, Appl
33	149	2.1	29	4	US-09-302-812-12	Sequence 12, Appl
34	149	2.1	29	4	US-09-511-477-12	Sequence 12, Appl
35	149	2.1	29	4	US-09-511-507-12	Sequence 12, Appl
36	149	2.1	427	4	US-09-134-001C-5143	Sequence 5143, Ap
37	148.5	2.1	1596	4	US-08-978-277A-4	Sequence 4, Appli
38	148	2.1	3248	1	US-08-353-700-1	Sequence 1, Appli
39	148	2.1	3248	5	PCT-US95-16216-1	Sequence 1, Appli
40	147	2.0	754	4	US-09-976-594-375	Sequence 375, App
41	146.5	2.0	2125	4	US-09-919-172-29	Sequence 29, Appl
42	146	2.0	1232	4	US-09-592-054-2	Sequence 2, Appli
43	145.5	2.0	913	4	US-08-971-089-4	Sequence 4, Appli
44	145.5	2.0	2482	1	US-08-328-254-6	Sequence 6, Appli
45	144.5	2.0	1085	1	US-08-431-080-28	Sequence 28, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-302-812-4  
; Sequence 4, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-302-812-4

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 5190.00 Matches: 976  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 72.00% Indels: 0  
DB: 4 Gaps: 0

US-09-302-812-3 (1-4069) x US-09-302-812-4 (1-976)

OY 167 ATGAATGGGCCCCGGCTGTGTAACCTGCACCAAGCGCCGCTGGGCGCGCTACA 226



1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20  
 227 ACTTCGGCGGTCTTCGGACGCCGGAGCTTTCCGAGACGACGAGCGCGTCTCGAC 286  
 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp 40  
 287 CCCAAGACGCTCAGTCGACGTTCAGGTCCACGCTCCCTCGCAGCGCTCGTCCACGGG 346  
 41 ProLysAspAlaHisValGlnPheArgValProSerSerProAlaCysValProGly 60  
 347 CAGCGCGGACAGCAGACGAGGCGGCCACCTCGCTTGTCTTCAACAAAAAGACTATTACC 406  
 61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80  
 407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAATTTGGATAGTAAGAA 466  
 81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100  
 467 AACAAACATACAAGAATAGAAATCATGATGATCTGTGTACAAAAGATAACTTTTACCAA 526  
 101 AsnAsnAsnThrArgLysGluSerMetSerSerValGlnLysAspAspPheTyrGln 120  
 527 CATTAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586  
 121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140  
 587 AAAAGTACACAGATTTTGAACCCAGCATCAGATCCAGCAATGTTAAGTGGCAAAATGAA 646  
 141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160  
 647 GGGAAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAAAACAGTAAACCTGGTACACAG 706  
 161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180  
 707 CAGTTTATGATGCTAACTATGATCGGTACCTCAAAATGATGATCATCAGTGACACAGAT 766  
 181 GlnPheSerAsnAlaAsnLysAspArgSerProGlnAsnAspHisSerAspThrAsp 200  
 767 AGTGAAGAGATAGACAAATCAACAGTCTTCTCAACAGTCTGAAGCTTGAAGTGAAG 826  
 201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220  
 827 CAGACTACGGAAGATGAACCGCCAGACAGAGCCAAAGCCACCAAGTGCAGCAAGTCT 886  
 221 GlnThrThrGluAspGlnHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240  
 887 TGCCATCTCTGGGAGACTGTGCAAGTTGTCCAGCAAGTGAAGTACGCTGGTGCCTCAAG 946  
 241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValProLys 260  
 947 AGTCCATTGTTCAGATGTTGCTCTGAGGATGTTGGTACTCGGTCAAAAATGACACAAA 1006  
 261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280  
 1007 TTGATTAGACAAAGAAATTCCTAGGAAATTCCTCCATTGAGAGGAAAGTGAACCC 1066  
 281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGluPro 300  
 1067 GAATCACCGATGATGGATTAATCTTAAATAATAGTTGTCAAGACTCAGAAAGCAGATGAG 1126  
 301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320  
 1127 GAGCAAGTCCAGGTTTGTATGAACAAAGAGATGGTACTCGGTCAAAAATGACAAATAAA 1186  
 321 GluThrSerProGlyPheAspGluGlnGluAspLysSerSerSerGlnThrAlaAsnLys 340  
 1187 CCTTCAAGGTTTCCAGCAAGACGCTGACATTTGAATTTAGGAAACGGTACTCTACTAAG 1246  
 341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360  
 1247 GCGGCTGAAGTTAGATTACATTTTCCAAATTTGAAGGAGGAGATGCGCATGGAAATGAT 1306

361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGluSerArgThrGlyMetAsn 380  
 1307 GATTAAATGCTAAACTACTGGAATATTTCTAGCCTGTAATGTAGATGCAGAAATTCCT 1366  
 381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400  
 1367 AAGCAACATCGAAAAAGGATCTTAAATACACAGATCATTTGATGACATGCCCAAGCA 1426  
 401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420  
 1427 GAGCACAGAAAGAAAGAACAGTGGAAACCAACATCAAAAGAACAGAAAGAAATCCCT 1486  
 421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440  
 1487 AAATACGTTCCACTCACCTTTCTCCAGATAAGAGTGGCTTGGAACTCCCATTTGAGGAG 1546  
 441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460  
 1547 ATGAGAAGAAATGCTCGTGTGGGATCCGCTCTCTCTTGAGACCATCTGCCAATCAC 1606  
 461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480  
 1607 ACAGTAACATTTCCGGTAGATCTTTTCGAGACGAGGAGAGTTCTCTAAACCTTTTCCAACA 1666  
 481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500  
 1667 CATTATAAGATTTGGGATTAACAGCATGTTAAATATGCTTGTTCAGAAACAAATTTG 1726  
 501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520  
 1727 TACCCAGTGAACATGAGAAATGTCGCGGAGCGGTGGAGAGTCTATTTCAG 1786  
 521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540  
 1787 ACTGCATCTTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAAATAC 1846  
 541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560  
 1847 AATGTGGCATTTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGATAAGGTA 1906  
 561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580  
 1907 CTTGAAGAAGCAGAACTCAACATTTATATCATCTCATCTTGTCTGATATGTGTGAAAT 1966  
 581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600  
 1967 GCATCTGTCTGCCAATATTTGCACCCAGCCAAATACCATCTCTGAAACAGAGATGAAT 2026  
 601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620  
 2027 CATTCCATCACATGTCGAGGAGACAGATTCAGCTTTTCTAGCTAATGCTTTCTCTGTC 2086  
 621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640  
 2087 ACATTTCCAGCAAGAAATCTAAGATGAATCGAGTATTCTAGTTTACCAGACATTAAC 2146  
 641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660  
 2147 TTCATTCGATTGTTTGAAGGACGTTTCATCAGAAACCGGAGAAACTTAAACGCTCTTC 2206  
 661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLysLysThrLeuPhe 680  
 2207 TGCTACTTTAGAGAGTACAGAGAAACCTTACTGGGTGGTGCATTTTACCAACACAG 2266  
 681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700  
 2267 AGCTTTGAAGATTTTCCAGATGGGAAGATGTGAAAAACCTTGTGACAGATTCGATGTC 2326  
 701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720  
 2327 ACTTAGAAGGTACCATAGAAAGAAATGCGCAAGGCATCTACAGTGGATTTTCAAAAT 2386  
 721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740



Db 301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320  
 QY 1127 GAGACAAGTCCAGGTTTGTATGAACAGAGATGGTAGTTCTCCCAACAGCAATAAA 1186  
 Db 321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsnLys 340  
 QY 1187 CCTTCAAGGTTCCAGCAAGAGCGCTGACATCAATTTAGGAAAGCGTACTCTACTAAG 1246  
 Db 341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360  
 QY 1247 GCGGCTGAAGTATGATTTACATTTCCAAATTTGAAGGAGAGAGAGTGCACCTGAAATGAAT 1306  
 Db 361 GlyGlyGluValArgLysHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380  
 QY 1307 GATTTAAATCTAAACTACCTGGAATATTTCTAGCCTGAATGTAGATGCGAATTTCT 1366  
 Db 381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400  
 QY 1367 AAGCAACATCGAAGAAAGGATTTCAAATCACAGATCATTTGATGACCTGCCCAAGCA 1426  
 Db 401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420  
 QY 1427 GAGCACAAGAAAGAAAGAGTGGAAACCAACATCAAAAGAACAGAAAGAAAGATCCCT 1486  
 Db 421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440  
 QY 1487 AATAAGTCTCCACTACCTTTCTCCAGATAAGAGTGGCTTGGAACTCCCATGAGGAG 1546  
 Db 441 LysTyrValProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460  
 QY 1547 ATGAGAAGATGCTCGCTGGATCGGCTCGCTCTGTGACCATCTGCCATCAC 1606  
 Db 461 MetArgArgMetProArgCysGlyLysLeuArgLeuProLeuLeuArgProSerAlaAsnHis 480  
 QY 1607 ACAGTAACTATTCGGTGTAGATCTTTTCGAGCAGAGAGTTCCTAAACCTTTTCAACA 1666  
 Db 481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500  
 QY 1667 CATTATAAGATTGTGGGTAACAACATGTTAAATGCTTGTTCAGAACAAATTTG 1726  
 Db 501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520  
 QY 1727 TACCACTGAGATGAGATGAGTGGAGTGGAGTGGGAGCGGCTGGAGCTCATTCAG 1786  
 Db 521 TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540  
 QY 1787 ACTGCACCTCTCAACAAATTTACACGACCCCAAACTTGAAGATGCTATTCTGAATAC 1846  
 Db 541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560  
 QY 1847 AATGGCATATTCTAAGAAATGGAGCTTTACAGCTTTGATTCGATTTCTGGGATAAGGTA 1906  
 Db 561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580  
 QY 1907 CTTCAAGACGAGAGTCAACATTTATATCAGTCCATCTTTCCTGATATGTTGCAAAAT 1966  
 Db 581 LeuGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600  
 QY 1967 GCATCTGTCTGCCAAATATTGCAACCCAGCAATACCACTCTCTGAAACAGAGATGAAT 2026  
 Db 601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620  
 QY 2027 CATTCCATCAATGTGCGAGAACAGATTCGAGTCTTTTAGCTAATGCTTTCTTCTGCG 2086  
 Db 621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPheCys 640  
 QY 2087 ACATTTCCACGACCAATGCTAAGTAAATGAAATCGAGTATTTAGTACCCAGACATTAAC 2146  
 Db 641 ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn 660  
 QY 2147 TTCAATCATGTTGTGAGGACGTTTCATCAAGGAAACCGAGAAATTTAAACGCTCTTC 2206  
 Db 661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680

QY 2207 TGCTACTTTAGAGAGTCAACAGAAAAACCTACTGGGTTGGTGACATTTTCAAGACAG 2266  
 Db 681 CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700  
 QY 2267 AGTCTTGAAGATTTTCCAGATGGGAAAGATGTGAAAAACCCCTTGACAGATTCGATGTC 2326  
 Db 701 SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal 720  
 QY 2327 ACTTACGAAGGTACCATAGAAAGAAATGCGCAAGGATGCTACAGGTGATTTTGGAAAT 2386  
 Db 721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740  
 QY 2387 CGTCTTGTGGGCTGGTGTAAACCTGACGAGCTGTGCAAGAAATCCGCTTTTGA 2446  
 Db 741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPheLeu 760  
 QY 2447 ATCAATCTCGAGTTGATTTATTTTCCCGCTCTTCACTGAGGTGCTGGATCACAATGATGT 2506  
 Db 761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
 QY 2507 CTAATTATCAGGTACTGACGAGTACAGTGAATACAGAGGTATGCTGAGACATATCGT 2566  
 Db 781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800  
 QY 2567 TGCTCCCGGAGCCACCAAGATGGAGTGAAGGAGCAGCTCGAGCGGGCTGCATCTGAG 2626  
 Db 801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820  
 QY 2627 ATCGTTTCCCATCATCATCTCTTCACTTCAAGACGCTACTCGATCAGTTTGTCCCTGAGAAA 2686  
 Db 821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840  
 QY 2687 ATGAGCGGAGCTGAAAGGCTTACTGTGATTTCTCGTCTCGTCTCGAGTTTCTTCAGAG 2746  
 Db 841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860  
 QY 2747 AATCTTCTCGAGTGGCCACGAGAACTGGGGCTGTGGCTTGGGGGTGATGCCAGG 2806  
 Db 861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880  
 QY 2807 TTAAGACCTTTAATACAGATATTGGCAGCTGCTGCGAGCTGAGCGAGATGTGTTATTTC 2866  
 Db 881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyrPhe 900  
 QY 2867 ACCTTTGGGAGCTCAGATTTGATGAGACATTTACAGCATGACATTTTCTTACTGAA 2926  
 Db 901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920  
 QY 2927 AGCAAACTCACTGCTGGAGATGTATAGCTCTTGTCTAGTACTACATGAAATGAGATGC 2986  
 Db 921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluCys 940  
 QY 2987 AGAACTGTTTCCACCCCTGGACCAAGATCAACAGCTTTATCCATTCATATACCATGCTGTC 3046  
 Db 941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960  
 QY 3047 GAGTCTGTGTCAGACCGCTGACCATTCAGGCAAGGACGAGGACC 3094  
 Db 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

RESULT 3

US-09-511-507-4  
 ; Sequence 4, Application US/095111507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPOLYMERASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH

```

; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-507-4

Alignment Scores:
Pred. No.: 0 Length: 976
Score: 5190.00 Matches: 976
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.00% Indels: 0
DB: 4 Gaps: 0

US-09-302-812-3 (1-4069) x US-09-511-507-4 (1-976)
QY 167 ATGAATGCGGCGCGCTGTGAACCTGCAACAAAGCGACCGCTGGGCGCGCTACA 226
DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20
QY 227 ACTTCGCGGCTGCTTCGGAGCCCGGAGCTTCCGAGCAGCGAGCGGCTCTCGAC 286
DB 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp 40
QY 287 CCAAGGAGCTCAGTCAGTCAGGTCACCGTCCACGCTCTCCGAGCTCGTCCAGGG 346
DB 41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60
QY 347 CAGGCGGACAGCAGCAGGAGCGCCACCTCGCTGTTTCAAAACAAAGACTATTACC 406
DB 61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrileThr 80
QY 407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATGAAGAA 466
DB 81 SerTrpMetAspThrLysGlyLysThrAlaGluSerGluSerLeuAspSerLysGlu 100
QY 467 AACACATACAGATAGATCCATGATGAGTCTGTACAAAAGATACTTTTACCACAA 526
DB 101 AsnAsnAsnThrArgileGluSerMetMetSerValGlnLysAspAsnPheTrpGln 120
QY 527 CATATGTAGAAAAATTAGTAAATGTTCTCAGCTAAGTCTTGATAAGTCACTCACTGAA 586
DB 121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140
QY 587 AAAAGTACACAGTATTGAACAGCATCAGACTGAGCAGCAATGTGTAAAGTGGCAAAATGAA 646
DB 141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160
QY 647 GGGAAACACAGGACAGCTTTGAAAGTGAACCTCAACAGTAAACCTGTGTACAGAG 706
DB 161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180
QY 707 CAGTTTAGTAAATGCTAAACATTGATCGGTCCACCTCAAAATGATGATCAGAGTACACAGAT 766
DB 181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200
QY 767 AGTGAAGAGATAGACAGCATCAACAGTTTCTCAACCTGTAAAGCTTGCAAATCCAAAG 826
DB 201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220
QY 827 CAGACTACGAGAGATCAACCCAGAGAGAGCCAAAGCCACCAAGTGCAGCAAGTCT 886
DB 221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240
QY 887 TGCCATCTCTGGGGAAGACTGTGCAAGTTGTTCAGCAAGATGAGATAGAGCTGTGCAAG 946

```

```

DB 241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260
QY 947 AGTCCATTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAACAA 1006
DB 261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280
QY 1007 TTGATTAGACAAGAAAGTTGCTAGGAATTTCTCTCCATTTCAGAGGAAGTGAAGTCAACC 1066
DB 281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300
QY 1067 GAATCACCAGTGGATGTGGATAATTCTAAATAATAGTTCTCAAGACTCAGAAGCAGATGAG 1126
DB 301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320
QY 1127 GAGACAAGTCAGGTTTGTGAACAAGAGATGTAGTCTCTCCCAACAGCAATATAA 1186
DB 321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340
QY 1187 CCTTCAGGTTCCAGCAAGAGACGCTGACATTCGAATTTAGAAACGGTACTCTACTAAG 1246
DB 341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360
QY 1247 GCGCGTGAAGTTAGATTACATTTCCAATTTCGAAGGAGAGAGTCCGACACTGGAATGAT 1306
DB 361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380
QY 1307 GATTAAATGCTAACTACCTGGAAATATTTCTAGCTTGAATGTAGAATGCGAAATTC 1366
DB 381 AspLeuAsnAlaLysLeuProGlyAsnIleSerLeuAsnValGluCysArgAsnSer 400
QY 1367 AAGCAACATGAAAGGATTTCTAAATACAGATCATTTGATGAGACTGCCAAAGCA 1426
DB 401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLysProLysAla 420
QY 1427 GAGGACAGAAAGAACACAGTGGGAAACCAACATCAAGAACAGAAAGAAAGATCCCT 1486
DB 421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440
QY 1487 AAATACGTTCCACCTCACCTTTCTCCAGATGAAGTGGTGGACCTCCATCCATTCAGAG 1546
DB 441 LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460
QY 1547 ATGAGAAGATGCTCGGTGGGATCCGCTCTCTGAGACCATCTGCCAATCAC 1606
DB 461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480
QY 1607 ACAGTAACCTATTCCGGTAGATCTTTTTCGAGCAGAGAGTTCCTTAAACCTTTTCCACA 1666
DB 481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500
QY 1667 CATTATAAGATTTGTGGGATAACAAGCATGTTAAATGCCCTTGTTCAGACAAATTTG 1726
DB 501 HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520
QY 1727 TACCAGTGGAGAGTGAATGGTGAGCAACTCGGGGAGCGGGAGGAGTCTATTCTCAG 1786
DB 521 TyrProValGluAspGluAsnGlyGlyArgThrAlaGlySerArgTrpGluLeuIleGln 540
QY 1787 ACTGCATCTTCAACAAATTTTACAGCCCCCAACCTTGAAGAGTGTCTATTCTGAAATAC 1846
DB 541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr 560
QY 1847 RATGTGCGCATATTCTAAGAAATGGACTTACAGCTTGTGATCGATTTCTGGGATAAGTA 1906
DB 561 AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580
QY 1907 CTTCAAGAGCAGAGACTCAACATTTATCAGTCCCATCTTCCCTGTGATGATGCTGAAATTT 1966
DB 581 LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle 600
QY 1967 GCATCTGTCTGCCAAATATTGTGACCCAGCCCAATACCACTCTCTGGAACAGAGATGAT 2026
DB 601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620

```

QY 2027 CATTCATCAATGTGCGAGCAACAGATGCGAGTCTTTTACGTAATGCTTTCTCTGC 2086  
 Db |||||  
 QY 621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640  
 Db |||||  
 QY 2087 ACATTTCCAGCAAGATGCTAGATGAATCCGAGTATCTAGTATCTACCCAGACATTAAC 2146  
 Db |||||  
 QY 641 ThrPheProArgAsnAlaLysMetLysSerSerGluTyLysSerTyLysProAspIleAsn 660  
 Db |||||  
 QY 2147 TTCAATCGATTGTTGAGGACGCTTCATCAAGAAACCGAGAACTTAAACCGCTCTTC 2206  
 Db |||||  
 QY 661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680  
 Db |||||  
 QY 2207 TGCTACTTTAGAAAGCTACAGAGAAACCTACTGGGTGGTGATCATTTACAAGACAG 2266  
 Db |||||  
 QY 681 CysTyLysPheArgValThrGluLysLysProThrThrGlyLeuValThrPheThrArgGln 700  
 Db |||||  
 QY 2267 AGCTTTGAGATTTCAGAAATGGGAAGATGTGAAACCCCTGACAGATTCGATGTC 2326  
 Db |||||  
 QY 701 SerLeuGluAspPheProGluThrGluArgCysGluLysProLeuThrArgLeuHisVal 720  
 Db |||||  
 QY 2327 ACTTAGAAGTACCATAGAAAGAAATGGCCAAAGCATGCTACAGGTGGATTTTGCAAT 2386  
 Db |||||  
 QY 721 ThrTyLysGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740  
 Db |||||  
 QY 2387 CGTTTTGTTGAGGTGTGTAACAGTGCAGGACTTGTGCAAGAAATCCGCTTTTAA 2446  
 Db |||||  
 QY 741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluLysLysArgPheLeu 760  
 Db |||||  
 QY 2447 ATCAATCCCTGAGTTGATTATTCACGCTCTTCACTGAGGTGCTGATCAATGATGT 2506  
 Db |||||  
 QY 761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
 Db |||||  
 QY 2507 CTAATTTATCACAGTACTGACAGTAGTACAGTGAATACACAGCTATGCTGAGACATATCGT 2566  
 Db |||||  
 QY 781 LeuIleIleThrGlyThrGluGlnTyLysSerGluTyLysThrGlyValAlaGluThrTyArg 800  
 Db |||||  
 QY 2567 TGCTCCCGAGCCACGAAGATGGAGTGAAGGAGCAGTCCAGCGCGCTGCATCAG 2626  
 Db |||||  
 QY 801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820  
 Db |||||  
 QY 2627 ATGTTTGCATCATGCTCTTCACTTCAGAGCTTACTCGATCAGTTGCTGCTGAGAA 2686  
 Db |||||  
 QY 821 IleValAlaIleAspAlaLeuHisPheArgArgTyLysLeuAspGlnPheValProGluLys 840  
 Db |||||  
 QY 2687 ATCAGACGCGAGCTGAACAGGCTTACTGTGAGTTTCTCGCTCGTGGAGTTTCTTCAGAG 2746  
 Db |||||  
 QY 841 MetArgArgGluLeuAsnLysAlaTyLysGlyPheLeuArgProGlyValSerSerGlu 860  
 Db |||||  
 QY 2747 AATCTTTCTGAGTGGCCACAGAAACTGGGGCTGTGTGCTCTTGGGGTGATGCCAGG 2806  
 Db |||||  
 QY 861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg 880  
 Db |||||  
 QY 2807 TTAAGACCTTAATACAGATATTGGCAGCTGCTGAGCTGAGGAGATGGTTTATTTTC 2866  
 Db |||||  
 QY 881 LeuLysAlaLeuIleGlnIleLeuAlaIleAlaIleAlaIleAlaIleAlaIleValTyPhe 900  
 Db |||||  
 QY 2867 ACCTTTGGGAGCTCAGAATTGATGAGACATTTTACAGCATGCAATTTTCCTTACTGAA 2926  
 Db |||||  
 QY 901 ThrPheGlyAspSerGluLeuMetArgAspIleTyLysSerMetHisIlePheLeuThrGlu 920  
 Db |||||  
 QY 2927 AGAAACTCACTGTTGAGATGTGATAGCTGTGTGCTAGTACTACATGATGAATGC 2986  
 Db |||||  
 QY 921 ArgLysLeuThrValGlyAspValTyLysLeuLeuLeuArgTyLysAsnGluGluCys 940  
 Db |||||  
 QY 2987 AGAACTGTTCCACCTCGACAGACATCAAGCTTTTATCATTTATATACCATGCTGTC 3046  
 Db |||||  
 QY 941 ArgAsnCysSerThrProGlyProAspIleLysLeuLysLeuTyLysProPheIleTyHisAlaVal 960  
 Db |||||  
 QY 3047 GAGTCTCTGTCAGACCGCTGACCATTCAGGGCAAGAGCAGGACC 3094  
 Db |||||  
 QY 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
 Db |||||

RESULT 4  
 US-09-302-812-2  
 ; Sequence 2, Application US/09302812B  
 ; Patent No. 6333148  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; EARLIER FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 ; US-09-302-812-2  
 Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 4635.50 Matches: 874  
 Percent Similarity: 92.83% Conservative: 32  
 Best Local Similarity: 89.55% Mismatches: 69  
 Query Match: 64.31% Indels: 1  
 DB: Gaps: 1  
 US-09-302-812-3 (1-4069) x US-09-302-812-2 (1-977)  
 QY 167 ATGAATGCGGGCCCGCTGTGAACCTCGCACCAAGCGACCGCTGGCGCGCTACA 226  
 Db |||||  
 QY 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 Db |||||  
 QY 227 ACTTCG---CCGGCTGCTTCGAGCGCCCGAGCTTTCGAGCAGGAGGAGCGCGCTCCT 283  
 Db |||||  
 QY 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 Db |||||  
 QY 284 GACCCCAAGAGCGCTCACGTGAGTTCAGGTCCACCGCTCCGCGCAGCGCTCCCA 343  
 Db |||||  
 QY 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
 Db |||||  
 QY 344 GGCAGCGCGGACAGCACAGAGCGCGCACCTCGCTGTTTTCACAAAGAGACTATT 403  
 Db |||||  
 QY 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 Db |||||  
 QY 404 ACCAGTTGGATGGACACTAAAGAAATCAAGACAGCGGAATCAGAAAGTTGGATAGTAA 463  
 Db |||||  
 QY 81 ThrSerTrpMetAspThrLysGlyLysThrValGluSerGluSerLeuHisSerLys 100  
 Db |||||  
 QY 464 GAAACCAACAATACAGATAGATCCATGATGAGTCTCTGACAAAGAGACTTTTAC 523  
 Db |||||  
 QY 101 GluAsnAsnThrArgGluGluSerMetSerValGlnLysAspAsnPheTyL 120  
 Db |||||  
 QY 524 CAACATAATGTAGAAAAATAGTAAATGTTTCTCAGCTAAGTCTTGTATAGTCACTCACT 593  
 Db |||||  
 QY 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 Db |||||  
 QY 584 GAAAGAGTACACAGTATTGTAACCGATCAGATCAGTCAGCAATGTGTAGTGGCAAAAT 643  
 Db |||||  
 QY 141 GluLysGlyThrGlnTyLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 Db |||||  
 QY 644 GAAGGAAACACACGAGCAGCTTTTGGAACTGAGTCAACAGTCAACCTGGTACCA 703  
 Db |||||  
 QY 161 GluGlyProHisSerGluArgLeuGluSerGluProAlaValThrLeuValPro 180  
 Db |||||  
 QY 704 GAGCAGTTTGTAGTAATGCTAACTTGTATCGTCAACCTCAACCTCAACCTGATCATCAGTGCACA 763  
 Db |||||











Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
Qy 404 ACCAGTTGGATGACACATAAGGAATCAAGACAGCGGAATCAGAAAGATTGGATAGTAAA 463  
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
Qy 464 GAAACCAACAATACAAAGATAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTTAC 523  
Db 101 GluAsnAsnAsnThrArgGluGluSerMetMetSerValGlnLysAspAsnPheTyr 120  
Qy 524 CAACATAATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGCTTGTGATAGTCACTCACT 583  
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
Qy 584 GAAAAAGTACACAGTATTTCAACCCAGCATCAGACTGCAGCAATGTTGTAAGTGGCAAAAT 643  
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
Qy 644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTGGTACCA 703  
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro 180  
Qy 704 GAGCAGTTAGTAATGCTAAATGATCGTCACTCACTCAAAATGATGATCAGATCAGACACA 763  
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
Qy 764 GATAGTGAAGAAATAGACAAATCAACAGTTTCTCACAACCTGTAAAGCTTGCAAAATGCA 823  
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
Qy 824 AAGCAGACTACCGAAGATGAACACGCCAGAGAGCCAAAGCCACAGAAAGTGCAGCAAG 883  
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
Qy 884 TCTTGCCATCTCGGGGAAGACTGTGCAAGTTGTGCAGCAAGATGAGATGACAGCTGGTGC 943  
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGlnGlnThrAspValValSer 260  
Qy 944 AAGAGTCCATGTCAGATGTGCTCTGAGAGTGTGGTACTGGGTCAAAAATGACAAC 1003  
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
Qy 1004 AAATTGATTAGACAAGAAAGTTCCTAGGAAATTCCTCCATTTGAGAAAGAAAGTGA 1063  
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu 300  
Qy 1064 CCCGAATCACCGATGAGTGGATTAATCTTAAATAAGTTGTCAAGACTCAGAAAGCAGAT 1123  
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
Qy 1124 GAGGAGCAAGTCCAGGTTTGTATGAACAAGATGTTAGTTCTCCCTCCCAACAGCAAAAT 1183  
Db 321 GluGlnThrSerProGlyPheAspGluGlnGluAspSerSerAlaGlnThrAlaAsn 340  
Qy 1184 AAACTTTCAAGGTTCCAAAGCAGAGAGCTGACATTTGAATTTAGAAACGTTACTTACT 1243  
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
Qy 1244 AAGGCGGTGAAGTTAGTTACATTTCCAAATTTCAAGGAGGAGAGAGTCCGACTGGGAATG 1303  
Db 361 LysGlyGlyIleArgLeuHisPheGlnPheGlnGlyGlyGluSerArgAlaGlyMet 380  
Qy 1304 AATGATTAAATGCTAAACTACCTCGAAATATTTCTAGCCTGAATGTAGATGCAAAAT 1363  
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
Qy 1364 TCTAAGCAACATGGAATAAAGATTTCTAAATCACAGATCATTTGATCAGACTGCCAAA 1423  
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
Qy 1424 CAGAGGACAGAAAGAAAGACAGTGGGAAACCAACATCAAGAACAGAAAGAGATC 1483  
Db 421 AlaGluAspLysArgLysGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440

Qy 1484 CCTAAATACGTTCCACTCACCTCTTCTCCAGATAAGAGTGGTGGAACTCCCATTTGAG 1543  
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
Qy 1544 GAGATGAGAAGAAATGCTCGGTGGGATCCGGCTGCTCTCTTGAGACCATCTGCCAAT 1603  
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn 480  
Qy 1604 CACAGTAACTATTCGGGTAGATCTTTTCGAGCAGAGAGAGTTCCTAAACCTTTTCCA 1663  
Db 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
Qy 1664 ACATTTATAAGATTTGTGGGATAACACAGCATGTTAAATGCTTAAATGCTTGTTCAGACAAAT 1723  
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
Qy 1724 TTGTACCCAGTGAAGATGAGATGCTGAGCGAATCGGGGAGCGTGGGAGCTCAT 1783  
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
Qy 1784 CAGACTGCACCTCTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAAA 1843  
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
Qy 1844 TACATGTGGCATATTTCTAAGAAATGGCACTTACAGCTTTGATTCGATTCGGGATAG 1903  
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuLeuAspPheTrpAspLys 580  
Qy 1904 GTACTTGAAGAGCAAGAGCTCAACATTTATATCATCTGCTCCCTGATATGCTGAAA 1963  
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
Qy 1964 ATTGCACCTGCTCTGCCAAATATTTGCCACCCAGCAATACCACTCTCTGAAAACAGAGATG 2023  
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
Qy 2024 AATCATTTCCATCACAAATGTGCGAGCAAGATGCGAGTCTTTAGCTAATGCTTTCTTC 2083  
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
Qy 2084 TGCATTTTCCAGCAAGAAATGCTAAGATGAATCGAGTATTCTAGTTTACCCAGACATT 2143  
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerSerGluTyrSerSerTyrProAspIle 660  
Qy 2144 AACTTCAATCGATGTTTCAGGACGCTTCATCAAGGAAACCCGAGAAACCTTAAACGCTC 2203  
Db 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
Qy 2204 TTCTGCTACTTTAGAAAGATCAGAGAAATAAACCTTACTGGGTGGTGCATTTTACAAGA 2263  
Db 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
Qy 2264 CAGAGTCTGAAGATTTTCCAGAAATGGAAGATGTGMAAAACCTTGCACAGCATTCAT 2323  
Db 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
Qy 2324 GTCACTTACCAAGTACCATAGAGAAATAATGGCCAAAGCATGCTACAGGTGGATTTTGA 2383  
Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
Qy 2384 AATCGTTTGTGGAGTGGTGAACCAAGTGCAGGACTCTGCAAGAGAAATCCCGCTTT 2443  
Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
Qy 2444 TTAATCAATCTGAGTTGATTTTTCACGGCTCTTCTCAGGTGCTGGATCACAATGAA 2503  
Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
Qy 2504 TGTCTAATTATCACAGGTACTGACAGTACAGTCAATACACAGGCTATGCTGAGCATAT 2563  
Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800

```
QY 2564 CTTGGTCCCGAGCCAGAGATGGAGTGAAGAGGACGACTCGAGCGCGCTGCACT 2623
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820
QY 2624 GAGATCCTTGCATCGATGCTCTTCACTTCAGACGCTACCTCGATCAGTTTGTGCTCAG 2683
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
821 GluIleValAlaIleAspAlaLeuHisPheArgTrpGlyLeuAspGlnPheValProGlu 840
QY 2684 AAATGAGACGCGAGCTGAACAGGCTTACTGTGATTTCTCCGCTCGAGATTCTTCA 2743
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860
QY 2744 GAGATCCTTCTGAGTGGCCACAGGAACTGGGGCTGTGCTTGGGGGTGATGCC 2803
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880
QY 2804 AGTTAAAGCCCTTAATACAGATATTCGAGCTGCTGAGCTGAGCGAGATGTGTTAT 2863
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValTyr 900
QY 2864 TTCACCTTGGGACTCAGAAATGATGAGAGACATTTACAGATGCAATTTTCCTTACT 2923
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920
QY 2924 GAAAGAAACTCACTGTGGAGATGTATAGCTGTGTGATGATGATGATGATGATGATG 2983
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940
QY 2984 TGCAGAAACTGTTCACCCCTCGACAGACATCAAGCTTTTATCATATACCATGCT 3043
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960
QY 3044 GTCAGTCTCTGCGAGACCGCTGACCATTCAGGCAAGACAGCGG 3091
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

RESULT 7
US-09-302-812-6
; Sequence 6, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOPHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; EARLIER FILING DATE: 1999-04-30
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-302-812-6

Alignment Scores:
Pred. No.: 0 Length: 968
Score: 4395.00 Matches: 837
Percent Similarity: 90.99% Conservative: 52
Best Local Similarity: 85.67% Mismatches: 78
Query Match: 60.97% Indels: 10
DB: 4 Gaps: 6

US-09-302-812-3 (1-4069) x US-09-302-812-6 (1-968)
QY 167 ATGAATCGGGCGCGCTGTGAACCTCGACCAAGAGCGCGCTGGGGCGCGCT--- 223
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19
224 ACAACTTCGCGCGCTGCTTCGAGCGCGCGAGCTTTCGAGCAGCAGCAGCGCGCTC 283
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValIleu 39
284 GACCCCAAGACGCTCACGTGCAGTTCAGGCTCCACCGCTCCGCGAGCTCGCTGCCA 343
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
344 GCGCAGCGGAGCAGCAGCAGCAGCGCGCCACCTCGCTTCTTTTCAAAAGAACTATT 403
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79
404 ACCAGTTGGATGACACTAAAGCAATCAACACAGCGCAATCAGAAATCTGGATAGTAA 463
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96
464 GAAACCAACAAATACAGAAATAGATCCATGATGAGTTCGTACAAAAGATACTTTTAC 523
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
97 GluAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 116
524 CAACATAATGTACAAAATAGTAATGTTTTCAGCTAAGTCTTGTATAGTCACTCACT 583
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136
584 GAAAAAGTACACAGTATTGTAACACGATCATGCTGACGATGCGCAATGTGTAGTGGCA 643
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysLysTrpGlnAsn 156
644 GAAGGAAACACACGCGAGCGAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTGGTACCA 703
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176
704 GAGCAGTTTGTAAATGCTTAACATTGATCGTCACTCAAAATGATGATCATAGTACACA 763
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196
764 GATAGTGAAGAGAATAGACAAATCAACAGATTTCTCACACTGTAAAGCTTCAAAATGCA 823
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216
824 AAGCAGACTACGGAAGATGAACACGCCAGAGAGGCCAAAGCCACAGAAAGTGCACAA 883
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233
884 TCTTCCCATCTCGGGAAGACGTGCAAGTTTCTAGCAAGATGAGATAGACGTGGTGCCA 943
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValIleuPro 253
944 AAGAGTCCATTTGTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAC 1003
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273
1004 AAATGATTAGACAGAAAGTTGCTCCTAGGAAATCTCCTCCATTTGAGAAGAAAGTGA 1063
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293
1064 CCCGNAATCCCGATGATGATTAATTTCTAAATAGTTGTCAAGACTCAGAACGAGAT 1123
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313
1124 GAGGAGACAGTCCAGGTTTGTGTAACCAAGAGATGTTGCTCTCCCAACACAGCAAT 1183
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332
1184 AAACCTTCAAGGTTTCCAGCAGAGACGCTGACATTTGAATTTAGAAACGGTACTTACT 1243
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352
1244 AAGGCGGTGAGTTAGATTACATTTCCATTTTCAAGGAGGAGAGAGTCCGCTGGAATG 1303
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371
```

	QY	1304	AATGATTATAAGTCTAAACTACCTCGAATAATTTCTAGCCTCAATGTAGAATGCAGAAAT	1363
	DB	372	SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer	391
	QY	1364	TCTAAGCAACATGGAAAAAGATTCTTAATAATCACAGATCATTTTGATGAGATGCCCAA	1423
	DB	392	SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys	411
	QY	1424	GCAGAGGACAGAGAAGAACCAAGTCGGGAACAACAACATCAAGAACAGAAAAGGAATC	1483
	DB	412	SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle	431
	QY	1484	CCTAATACAGTTCACCTCCACCTTCCTCCAGATAAGAAGTCGCTGGAACTCCCAATTGAG	1543
	DB	432	ProLysTyrlleProProAsnLeuProProGluLysLysTrpLeuglyThrProileGlu	451
	QY	1544	GAGATGAGAAGATCGCTCGGTCTGGATCCGGCTGCCNCTCTTGAGACCATCTGCCAAT	1603
	DB	452	GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer	471
	QY	1604	CACACAGTAACTATTCGGGTAGATCTTTTTGCGAGCAGGAGAAGTTCCTAAACCTTTCCA	1663
	DB	472	HisThrValThrValargValaspLeuLeuArgAlaglyGluValProLysProPhePro	491
	QY	1664	ACACATTATAAAGATTGTGGGATAACAGACATGTTAAATGCTTGTTCAGACAAAAT	1723
	DB	492	ThrHisTyrllysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	511
	QY	1724	TTGTACCCAGTAGACATGAGTAGTGAGCAGCAACTGCGGGAGCCGGTGGGAGCTCAAT	1783
	DB	512	LeuTyProValGluaspGluAuengLygluarthrAlaglySerArgTrpGluLeulle	531
	QY	1784	CAGACTGCATCTCTCAACAAATTTACACGACCCCCAAACTTGAAGATGCTATTCTGAAA	1843
	DB	532	GlnThrAlaleuLeuAuenLysPheThrArgProGlnAsnLeuLysAspAlalleLeuLys	551
	QY	1844	TACAATGTGCATATTCTAAGAAATGGGACTTTACAGCTTTTGATCGATTTCTGGGATAG	1903
	DB	552	TyrAsnValAlatySerLysLysTrpAspPheThrAlaleuValAspPheTrpAspLys	571
	QY	1904	GTACTTGAACAAGCAGNAAGCTCAACATTTATATCAGTCCACTTGTGCTGATATGGTGAAA	1963
	DB	572	ValleuGluGluAlaGluAlaGlnHisLeuryrGlnSerlleLeuProaspMetValLys	591
	QY	1964	ATTGCACTCTGTGCGCAATATTTCACCCAGCCCAATACCACTCTCTGAAACAGAGATG	2023
	DB	592	IleAlaleuCysLeuProProAsnLleCysThrGlnProilleProLeuLeuLysGlnLysMet	611
	QY	2024	AATCAITTCATCACAATGTGCGAGGACAGATGCGCAGTCTTTTAGCTAATGCTTTCTTC	2083
	DB	612	AsnHisSerValThrMetSerGlnGlnGlnllealaSerleuLeuAlaasnAlaPhePhe	631
	QY	2084	TGCACATTTCCACGACCAAAATGCTTAAGATGAAATFCGGAGTATTTCTAGTTACCCAGACAT	2143
	DB	632	CysThrPheProArgAgaenaAlalyMetLysSerGluTyrrSerSerTyrrProAspile	651
	QY	2144	AACTTCAATCGATTGTTTGGAGCAGTTCATCAGGAAAACCGGAGAACTTAAACGCTC	2203
	DB	652	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu	671
	QY	2204	TTCTGCTACTTTTACAAGAGTCACAGAGAAAAAACTACTGCGTGGTGACATTTTACAAGA	2263
	DB	672	PheCysTyrrPheArAgavalthrGluLysLysProThrGlyLeuValThrPheThrArg	691
	QY	2264	CAGAGTCTTTGAAGATTTTCCAGAATGGGAAAGATGTGAAAACCCCTTGACACGATTCAT	2323
	DB	692	GlnSerLeuGluAuspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis	711
	QY	2324	GTCACTTACGAGGTACCATAGAGAAATGGCCMACGCATGCTACAGTGGATTTTGCA	2383
	DB	712	ValThrTyrrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla	731

## Alignment Scores:

Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: 4 Gaps: 6

US-09-302-812-3 (1-4069) x US-09-511-477-6 (1-968)

QY 167 ATGAATGCGGCGCCCGCTGTGACCTGTGACCAAGCGACCGCTGGGCGCGCT--- 223  
 DB 1 MccSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGCGCTGTTCGCGCGCGCGCTGTTCGCGCGCGCGCTGTTCGCGCGCGCT 283  
 DB 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 284 GACCCCAAGACCTCACGTGCGAGTTCAGGTCCTCCACCGCTCCGCGCTCCCA 343  
 DB 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
 QY 344 GGCACGCGCGGACACAGACGACGCGCGCTGTTCGCGCGCGCTGTTCGCGCGCTGTTC 403  
 DB 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 404 ACCAGTTGGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTGGATAGTAAA 463  
 DB 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 464 GAAACACAAATACAGAATAGATCCATGATGAGTTCGTACAAAAGATACCTTTTAC 523  
 DB 97 GluAsnAsnThrArgIleAspSerMetMetSerValGlnLysAspAsnPheTyr 116  
 QY 524 CAACATAATGTAGAAAATAGTAATGTTCTCAGCTAAGTCTGTAGTACCTCACT 583  
 DB 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLysSerProThr 136  
 QY 584 GAAAAAGTACACAGTATTGAAACAGCATCAGCTGCGCAATGTGTAGTGGCAAAAT 643  
 DB 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
 QY 644 GAAGGAAACACACGAGCAGCTTTGGAAGTGAACCTCAACAGTAACCTGTGACCA 703  
 DB 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 704 GAGCAGTTTAGTAATGCTAACATTCATGCTGCTCAGCTCAAAATGATGATCAGTACAC 763  
 DB 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 764 GATAGTGAAGAGAAATAGAGACAAATCAACAGTTTCTCACAACCTGTAAAGCTTCAAAATGCA 823  
 DB 197 AspHisGluGluAspArgAspAsnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 QY 824 AAGCAGATCAGGAGATCAACACCCAGAGAGCCAAAGCCACAGAGTGCAGCAG 883  
 DB 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233  
 QY 884 TCTTCCCATCTCTGGGAGACGTGCAAGTCTCAGCAAGATGAGATAGACTGTGTGCA 943  
 DB 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGlnGluValAspValLeuPro 253  
 QY 944 AAGAGTCCATTCAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATCACAAC 1003  
 DB 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1004 AAATGATTAGACAGAAAGTTGCTAGCAATTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1063  
 DB 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293  
 QY 1064 CCCGAATCACCGATGGATGATTAATCTAAATAATAGTTGTCAAGACTCAGAACAGAT 1123  
 DB 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

QY 1124 GAGGACAGCAAGTCCAGGTTTGTGATGAACAAGAAGTGTAGTTCTCTCCCAACAGCAAAAT 1183  
 DB 314 GluGlnThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332  
 QY 1184 AAACCTTCAAGGTTCCAAAGCAAGAGCGCTGACATTGAATTTAGGAAACCGTACTCTACT 1243  
 DB 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
 QY 1244 AAGGCGGTTGAAGTATGATTCATTTCCATTTGAAGGAGGAGAGAGTGCATCGCATGGAATG 1303  
 DB 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371  
 QY 1304 AATGATTTAAATGCTAAACTACCTCGAATATTTCTAGCTGATGTAGTATAGTATCAGAAAT 1363  
 DB 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 QY 1364 TCTAAGCAACATCGAAAAAGGATTTCTAAATCAGATCATTTGATGAGATCGCCCAA 1423  
 DB 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
 QY 1424 GCAGGAGCAGAGAAAGAAAGACAGTGGGAAACCAACATCAAGAAACAGAAAGAGATC 1483  
 DB 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 QY 1484 CCTAAATACGTTCCACCTCACCTTCTCCAGATAGAGTGGCTCGAACTCCCATGAG 1543  
 DB 432 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 451  
 QY 1544 GAGATGAGAGAAATCGCTCGGTGTGGGATCCGCGCTCTCTTGAGACCATCTGCGCAAT 1603  
 DB 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
 QY 1604 CACACAGTACTATTCGGGTAGATCTTTTGCAGACAGAGAGTCTCTAAACCTTTTCA 1663  
 DB 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
 QY 1664 ACACATTAATAAGATTTGTGGATACCAAGCATGTTAAATGCCTTGTTCAGAACAAAT 1723  
 DB 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 QY 1724 TTGTACCCAGTGAAGATGAGAAATGCTGAGCGAACTGCGGGGAGCGGTGGAGTCTCAT 1783  
 DB 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
 QY 1784 CAGACTGCATCTCAACAAATTTACAGACCCCAAACTTGAAGTCTATCTTGAA 1843  
 DB 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLys 551  
 QY 1844 TACATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTTGTATGATGATGATGATG 1903  
 DB 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
 QY 1904 GTACTTGAAGACAGACAGTCAACATTTATCATGCTCCATCTTCCCTGATATGTGTGAA 1963  
 DB 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
 QY 1964 ATTGCATCTCTGTGCCAAATATTGCAACCCAGCAATACCACTCTCTGAACAGAGATG 2023  
 DB 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
 QY 2024 AATCATTCATCAATGTCGAGGAAACAGATTCGAGTCTTTAGCTTAATGCTTTCTCTC 2083  
 DB 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 QY 2084 TGCACTTCCACAGCAGAAATGCTAAGATGAATTCGGAGTATTCCTAGTATCCAGACAT 2143  
 DB 632 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
 QY 2144 AACTTCAATCATTTGTTGAGGAGCTTCATCAAGGAAACCGAGAACTTAAACGCTC 2203  
 DB 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

```

QY 2204 TTCTGCTACTTTAGAGAGTCAAGAGAAACCTACTGGTGTGGATTTACAGA 2263
Db 672 PheCysTyrPheArgArgValThrGluGlySerProThrGlyLeuValThrPheThrArg 691
QY 2264 CAGAGTCTTCAAGATTTTCCAGATGGAAGATGTGAAAAACCTTTGACACGATTCAT 2323
Db 692 GlnSerLeuGluAspPheProGluTyrGluArgCysGluGlyProLeuThrArgLeuHis 711
QY 2324 GTCACTTACAGAGTACATAGAGAAATGCGCAAGCATGCTACAGGTGGATTTGCA 2383
Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
QY 2384 AATCGTTTTGTGAGGTGGTGAACAGTCCAGAGTCTTGCAAGAGAAATCCGCTTT 2443
Db 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluIleArgPhe 751
QY 2444 TTAATCAATCTGAGTTATTTCACGGCTCTTCACTGAGGTGCTGGATCAATGAA 2503
Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
QY 2504 TGTCTAATTATCACAGGTACTGACGAGTACAGTGAATACACAGGCTATGCTGACACATAT 2563
Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791
QY 2564 CGTTGGTCCCGAGCCACGAGATGGAGTGAAGAGGACGACTCGGAGCGCGCTGCAC 2623
Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluGlyAspAspTrpGlnArgArgCysThr 811
QY 2624 GAGATCGTGCATCGATGCTTCACTTCAGAGCTACCTCGATCAGTTGTGCTGAG 2683
Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831
QY 2684 AAAATGAGCGCGAGCTGAACAGGCTTACTGTGATTTCTCCGCTCGGAGTTCTTCA 2743
Db 832 LysValArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851
QY 2744 GAGAACTTTCTGAGTGGCCACAGGAACTCGGGCTGTGGTGTCTTTGGGGTGTAGGCC 2803
Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871
QY 2804 AGTTAAAGCTTAAATACAGATTTGGCAGCTGCTGAGCTGAGCGAGATGTGGTTAT 2863
Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaGluArgAspValValTyr 891
QY 2864 TTCACCTTGGGACTCAGATTCAGTACGACATTTACAGCATGTCACATTTCTCTTACT 2923
Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
QY 2924 GAAAGGAACTCACTGTGGAGATGTGTATAGCTGTGTGCTTACGATACATCAATGAAGAA 2983
Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931
QY 2984 TGCAGAACTGTTCCACCCCTGGACGACATCAAGCTTTTATCATTATATACCATGCT 3043
Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951
QY 3044 GTCCAGTCTCTGTCAGAGACCGCTGACCATTCAGGGCAAGAGGACGAGGCC 3094
Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

```

RESULT 9

```

US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201

```

```

; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

Alignment Scores:
Pred. No.: 0 Length: 968
Score: 4395.00 Matches: 837
Percent Similarity: 90.99% Conservative: 52
Best Local Similarity: 85.67% Mismatches: 78
Query Match: 60.97% Indels: 10
DB: 4 Gaps: 6

US-09-302-812-3 (1-4069) x US-09-511-507-6 (1-968)
QY 167 ATGAATGCGGCGCGCGCTGTGAACCTGCACCAAGCGACCGCTGGCGCGCT--- 223
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19
QY 224 ACAACTTCGCGCGCTGCTTCGAGACGCCGCGAGCTTCCGAGCAGCGCAGAGCGCGCTC 283
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39
QY 284 GACCCCAAGACGCTCACTGTCAGGTTCAGGGTCCCGCTCCCGCTCCGCGCGCTGGTCCCA 343
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
QY 344 GCGCAGCGCGGACAGCAGCAGCGCGCGCTCGCTCGCTCCCGCTCCCGCTGGTCCCA 403
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79
QY 404 ACCAGTTGGATGACACTAAAGGAATCAACACAGCGGAATCAGAAAGTTTGGATAGTAAA 463
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96
QY 464 GAAACCAACAAATCAAGAATAGAAATCAATGAGTTCTGTACAAAAGATTAATTTTAC 523
Db 97 GluAsnAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 116
QY 524 CAACATATCTACAAAATAGTAAATGTTCTCAGCTAAGTCTTGAAGTCACTCACTCACT 583
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136
QY 584 GAAAAAGTACACAGTATTTGAACACGATCAGATCGCAGCAATGTGTAAAGTGGCAAAAT 643
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156
QY 644 GAGGGHAAACACACGAGGAGCTTTTGGAAAGTGAACCTCAACAGCTAAACCTGGTACCA 703
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176
QY 704 GAGCAGTTTATGATGCTTAACATTCGCTGCTCACTCACTCACTCACTCACTGACACA 763
Db 177 LysGlnLeuSerAsnAlaAsnIleGlnSerProHisThrAspAspHisSerAspThr 196
QY 764 GATAGTGAAGAGAAATAGACAATCAACAGTTTCTCAAACTGTAAAGCTTGGCAATGCA 823
Db 197 AspHisGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216
QY 824 AAGCAGACTACGGAAGATGAACACGCGCAGAGAGCCAAAGCCACCAAGAGTGCAGCAAG 883
Db 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233
QY 884 TCITGCCATCTCGGGAAGACTGTGCAAGTTGTTCAGCAAGATGAGATAGAGCTGGTGCCA 943
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253

```







QY 3011 GACATCAAGCTTTATCCATTATACCATGCTGTGAGTCTCTGAGTCTGTGAGAGCCGCTGAC 3070  
Db 582 LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAsp 598  
QY 3071 CATTCAGGGCAA---AGGACAGGAGCCTGAGGAGCCGAGCGAATAGCA 3115  
Db 599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerArgValAla 614  
RESULT 11  
US-09-511-477-8  
; Sequence 8, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; PRIOR FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-477-8  
Alignment Scores:  
Pred. No.: 4,06e-102 Length: 768  
Score: 1081.00 Matches: 246  
Percent Similarity: 57.89% Conservative: 99  
Best Local Similarity: 41.28% Mismatches: 197  
Query Match: 15.00% Indels: 54  
DB: 4 Gaps: 16  
US-09-302-812-3 (1-4069) x US-09-511-477-8 (1-768)  
QY 1412 AGACTCCCAAGCAGAGACAGAGAAAGACAGTGGAAACCAACATCAAGACA 1471  
Db 45 ArgMetSerLysProAspGlyGlyIleSerGluIleGluThrGluGlu----- 62  
QY 1472 GAAAGGAAGATCCCTAAATACGTTCACCTCCACCTC-----TCTCCAGATAAGAGTGG 1525  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74  
QY 1526 CTTGGAATCCCATTTAGAGATG-----AGAAGATGCTCGGTGGATCCGGCTG 1579  
Db 75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94  
QY 1580 CCTCTCTGAGACCATCTGCCAATCAACAGTAATCTTCCGGTGTAGATCTTTTGGCGACA 1639  
Db 95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114  
QY 1640 GGAGAAAGTTCCTAAACCTTTTCCACACATTAATAAGATTGTGGGATAACAGCATGTT 1699  
Db 115 -----ProProArgProTyrLysSerProGlyLys-----TrpAspSerGluHisVal 130  
QY 1700 AAAATGCTTTGTCAGACAAATTTGTACCCAGTGAAGATGAGATGGTGAAGCACT 1759  
Db 131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150  
QY 1760 GCGGGGCGGTGGAGCTCATTCAGACTGCTCTCTCAACAAATTTACACGCCCA 1819  
Db 151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170  
QY 1820 AACTTGAAGCATGCTATTCTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACA 1879  
Db 1879

Db 171 GluLeuGlnAlaAlaIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190  
QY 1880 GCTTTGATCGATTTCGGGATTAAGTACTTGAAGAAGCAGAGCTCAACATTATATCAG 1939  
Db 191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210  
QY 1940 TCATCTTCCTGATATGTTGAAATTCACCTCTGTCGCCAAATATTTGCACCCAGCCA 1999  
Db 211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230  
QY 2000 ATACCACTCCTGAAACAGAGATGAATCATTCATCAATGTCGAGGAGAACAGATTGCC 2059  
Db 231 ValProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250  
QY 2060 AGTCTTTAGCTAATGCTTTCTTCACATTTCCACGACGAAATGCT---AAGATGAAA 2116  
Db 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
QY 2117 TCGGAGTATTCTAGTTTACCCAGACATTAACTTCAATCGATTGTTTGAGGGACGTTTCATCA 2176  
Db 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
QY 2177 AGGAAACCGAGAACTTAAACGCTTCTGCTACTTTAGAGAGTC-----ACAGAG 2230  
Db 291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310  
QY 2231 AAAAAA-----CCTACTGGTGGTGGACATTTACAGACAGAGT-----CTT 2272  
Db 311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgSerGlyLeuPro 330  
QY 2273 GAAGATTTTCCAGATGGAAAGATGTGAAAAACCTTG-----ACAGATTGCAATGC 2326  
Db 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
QY 2327 ACTTACGAAGGTACCATAGAGAAATGCCAAGCATGCTACAGTGGATTGTTGCAAT 2386  
Db 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370  
QY 2387 CGTTTGTGGAGGTGTGTAAACAGTGCAGACTGTGCAAGAGAAATCCGCTTTTA 2446  
Db 371 LysTyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheVal 390  
QY 2447 ATCAATCTCGAGTGTGATTATTTACCGCTTTCACGTGAGTCTGTGATCAATGAATGT 2506  
Db 391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
QY 2507 CTAATTATCAGAGTACTGAGCAGTACAGTGAATACACAGGTATGCTGAGACATATCGT 2566  
Db 411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
QY 2567 TGTCTCCGAGCCACGAGATGGAGTGAAGAGGAGCAGTGGAGCGGCTGCACTGAG 2626  
Db 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450  
QY 2627 ATCGTTGCCATCGATCTCTTCACCTTCAGACGCTACCTCGATCAGTTTGTGCTGAGAA 2686  
Db 451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisIleGlnTyrArgGluAspLeu 470  
QY 2687 ATGAGACGGAGCTGAACGAGTCTTACTGTGATTTCTCCGT----- 2728  
Db 471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490  
QY 2729 CTTGGAGTTTCTTCAGAGATCTTTCTGCGAGTGGCCACAGGAAACTGGGGCTGTGGTGC 2788  
Db 491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502  
QY 2789 TTTGGGGGTGATGCCAGGTAAAAAGCCTTAATACAGATATTGCGAGCTGCTGAGCTGAG 2848  
Db 503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522  
QY 2849 CGAGATGTGGTATTATTCACCTTTGGGAGCTCAGAAATGATGAGACATTTACACATG 2908  
Db 523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542



QY 2909 CACATTTTCCTTACTGAAAGAAACTCACTGTTGGAGATGTGTATAAGCTGTTGCTACGA 2968  
Db 543 TrpLeuLeuPheArgAspGlyThrValGlnGlnLeuTrpSer---IleLeuArg 561  
QY 2969 TACTCAATGAAGATGAGCAAACTGTTCCACCCTGGACCA----- 3010  
Db 562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581  
QY 3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCAGTCTCTGTCAGAGACCGCTGAC 3070  
Db 582 LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValArgAsp 598  
QY 3071 CATTCAGGGCAA---AGCAGGAGACCTGAGGACCGGAGCAATAGCA 3115  
Db 599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerArgValAla 614

RESULT 12  
US-09-511-507-8  
; Sequence 8, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIA2 201  
; CURRENT APPLICATION NUMBER: US/09/511.507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 8  
; LENGTH: 768  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
US-09-511-507-8

Alignment Scores:  
Pred. No.: 4,06e-102 Length: 768  
Score: 1081.00 Matches: 246  
Percent Similarity: 57.89% Conservative: 99  
Best Local Similarity: 41.28% Mismatches: 197  
Query Match: 15.00% Indels: 54  
DB: 4 Gaps: 16

US-09-302-812-3 (1-4069) x US-09-511-507-8 (1-768)

QY 1412 AGACTGCCAAAGCAGGACGACAGAAAGAAAGCAGTGGGAAACCAACATCAAGAAACA 1471  
Db 45 ArgMetSerLysSerProAspGlyClyLeuSerGluIleGluThrGluGluGlu----- 62  
QY 1472 GAAGAGAGATCCCTAAATACGTTCACCTCACCTT-----TCTCCAGATAAGAGTGG 1525  
Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTip 74  
QY 1526 CTTGGAACCTCCATTGAGGAGATG-----AGAGAATGCTCGTGTGGGATCCGGCTG 1579  
Db 75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94  
QY 1580 CCTCTCTTGAGACCATCTGCGCAATCACACAGTAACATTCCTGGGTAGATCTTTTCGAGCA 1639  
Db 95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114  
QY 1640 GGAGAAGTTCCTAAACCTTTTCCAAACAATTATAAGATTGTGGGATPAAACAGCATGTT 1699  
Db 115 -----ProProArgProTyrLysSerProGlyLys-----TipAspSerGluHisVal 130  
QY 1700 AAAATGCCTTGTTCAGAACAAATTTGTACCAGCTGGAAGATGAGAAATGGTGAGCGACT 1759

131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150  
1760 GCGGGAGCGGCTGGAGCTCATTAGACTGCACTTCTCAACAATTTACAGCAGCCCAA 1819  
151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170  
1820 AACTTGAAGATGCTATTCTGAAATACAAATGTGGCATATTCTAAGAAATGGAGCTTTACA 1879  
171 GluLeuGlnAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190  
1880 GCTTTGATTCATTTCTGGGATAAGTACTTGAAGACAGCAAGCTCAACATTTATATACAG 1939  
191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu 210  
1940 TCCATCTTGCTGATATGTTGAAATTTGCACCTCTGCTGCCAAATATTTCACCCAGCCA 1999  
211 AspleuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230  
2000 ATACCCTCTGAAACAGAGATGATCATTCATCATCATCATGTGCGGAGCAAGATTGCC 2059  
231 ValProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250  
2060 AGTCTTTTAGCTAATGCTTTCTTCACATTTCCACGACGAAATGCT---AAGATCAA 2116  
251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
2117 TCGAGTATTCTAGTTTACCAGACATTAATCAATCGATTGTTTGGAGGACGTTTCACTCA 2176  
271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
2177 AGGAACCGGAGAACTTAAACGCTTCTGCTACTTACAGAGAGTC-----ACAGAG 2230  
291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310  
2231 AAAAAA-----CCTACTGGTGTGTCACATTTCAAGACAGAGT-----CTT 2272  
311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330  
2273 GAAGATTTTCCAGAAATGGGAAAGATGTGAAAAACCCCTTG-----ACACGATTCATGTC 2326  
331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
2327 ACTTACGAGGTACCATAGAGAAATGGCCAGGCATGCTACAGTGGATTTTGCAAT 2386  
351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuGlnValAspPheAlaAsn 370  
2387 CGTTTCTGAGGTGTGTAACCCAGTGCAGGACTTGTGCAAGAGAAATCCGCTTTTA 2446  
371 LysTyrLeuGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390  
2447 ATCAATCTGAGTTGATTATTTCACGCTCTTCACTGAGTGTCTGGATCACAATGATGT 2506  
391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
2507 CTAATTATCACAGTACTGAGCAGTACAGTGAATACACAGCTATGCTGAGACATATCGT 2566  
411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
2567 TGTGTCGCGGACCCAGCAAGATGGAGTGAAGGAGGACGACCTGCGAGCGCGCTGCAGT 2626  
431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyAspGlnThrAla 450  
2627 ATCGTTGCCATTCAGTGTCTTCTCATTTCAGAGCTACCTCGATCAGTTTGGCTGAGAAA 2686  
451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu 470  
2687 ATGAGACGCGAGCTGAACAGGCTTACTGTGATTTCTCGT----- 2728  
471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490  
2729 CCTGGAGTTCCTTCAGAGAATCTTTCTGAGTGGCCACAGAAACTCGGGCTGTGGTCC 2788

```
Db 491 ProGly-----ValalaThrGlyAsnTrpGlyCysGlyAla 502
QY 2789 TTTGGGGTGATGCCAGGTTAAAGCCTTAATACAGATATTGGCAGCTGCTGAG 2848
Db 503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522
QY 2849 CGAGATGTGTTTATTCACCTTTGGGACTCAGATTGATGAGACATTTACAGCATG 2908
Db 523 ArgProLeuAlaTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542
QY 2909 CACATTTTCTTACTCAAGGAACTCAGCTGTGGAGATGTGTATAGCTGTGTACGA 2968
Db 543 TrpLeuLeuPheArgAsnAspGlyThrValGlnGlnLeuTrpSer---IleLeuArg 561
QY 2969 TACTACAATGAGAATGCGAAGAACTTCCACCCCTGGACCA-----3010
Db 562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
QY 3011 GACATCAAGCTTTATCATTATACATATACCTGCTGTGCTGCTGTCAGAGCCGCTGAC 3070
Db 582 LysLysLysLeuTyrAspPheIle-----LysGluLeuLeuLysLysValArgAsp 598
QY 3071 CATTTCAGGGCAA---AGGACAGGGACCTCAGGAGCCGCGAGCAATAGCA 3115
Db 599 ValProGlyGluGlyAlaSerAlaGluAlaGlySerSerArgValAla 614

RESULT 13
US-09-302-812-10
; Sequence 10, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-302-812-10

Alignment Scores:
Pred. No.: 2,73e-38 Length: 726
Score: 463.00 Matches: 207
Percent Similarity: 38.11% Conservative: 123
Best Local Similarity: 23.90% Mismatches: 278
Query Match: 6.42% Indels: 258
DB: 4 Gaps: 37

US-09-302-812-3 (1-4069) x US-09-302-812-10 (1-726)
QY 599 TATTGACACCATCAGCTGAGCAATGTGTAACTGGCAAAATGAAGGAAACACAG 658
Db 27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46
QY 659 GAGCAGCTTTTGAAGTGAACCTCAACAGTAACTGCTGCTACAGCAGCAGCTTTAGTAAT 718
Db 47 GluSerLysGluAsp-----ProGluGluProLysSer 57
QY 719 GCTAACATTGATCGTCACTCAAAATGATGATCAGCTGACACAGATAGTGAAGAAT 778
Db 58 ArgAspValPheValSerGlnSerSerGlnSerGluSerGluGlnGluAspSerAlaGluAsn 77
```

```
QY 779 -----AGAGACATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
Db 78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97
QY 815 GCAAATGCAAGACAG-----ACTACGGAAGATGAACACACCCAGAGAGCAAA 862
Db 98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp-----115
QY 863 AGCCACCAAGATGCGCAAGTCTTCCCATCTGGGGAAGACTGTGCAAGTGTCTGACAA 922
Db 116 AsnHis-----LysSerThrGluProMetGluGlu-----125
QY 923 GATGAGATAGACGTGGTCCCAAGACTCCATTGTCCAGATGTTGGCTCTCAGAGATGTTGT 982
Db 126 -----AspValAsnAsnLysSerAsnIle-----AspValAla 136
QY 983 ACTGGTCAAAAAATGACAAATGATTAGACAGAAAAGTTGCTTAGGAAATTTCTCCT 1042
Db 137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn-----150
QY 1043 CCATTTGAGAAGAAAGTGAACCCGAATCACCAGTGTGATGTGATTAATTTCTAAAATAGT 1102
Db 151 -----AsnLysGluMetArg-----AspGlyGluGlnValGlnGlnLeu 163
QY 1103 TGTCAAGACTCAGAAGCAGATGAGGAG-----ACAAGTCCAGGTTTGTGATGAACAA 1153
Db 164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183
QY 1154 GAA-----GATGTAGTTCTCCCAACACAGCAAAATAAACCTTCAAGG 1195
Db 184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203
QY 1196 TTCCAAGCAAGAGACGCTGACATTTGATTT-----AGGAACCG 1234
Db 204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223
QY 1235 TACTCTACTTAAGGCG-----1249
Db 224 ValArgThrSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243
QY 1250 GGTGAAGTTTAGATTACAT-----TTCCAATTTGAAGGAGGA 1285
Db 244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263
QY 1286 GAGAGTCGCACTGGAATGATTAATGATTAATGCTAACTACTCTGGA-----1330
Db 264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282
QY 1331 -----AATATTTCTAGCCTGAATGTAGAAATGCGAAATTTCTAAGCAACATGGA 1378
Db 283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302
QY 1379 AAAAAGGATTTAAATACAGATCATTTGATGAGACTCCCAACAGCAGAGACAGAAAGA 1438
Db 303 ProLys-----IleValLeuPro-----GlnArg 310
QY 1439 AAGAACAGTGGGAACCAACATCAAGAACAGAA-----AGGAAG 1480
Db 311 TrpArgGluPheAspSerArgGlyArgArgAspSerTyrPheTyrPheLysArgLys 330
QY 1481 ATCCCTTAAATGCTCCACCTCCCTTCTCCAGATAAGAACTGCTTGGAACTCCCAT 1540
Db 331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350
QY 1541 GAGGAGATGAGAAGAATG---CCTCGTGTGGGATCCGCTCCTCTTGTGACCATCT 1597
Db 351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu-----368
QY 1598 GCCAATCAACAGTAATCTATTCGGGTAGATCTTTTCGAGCAGGAGAGAACTTCTTAAACCT 1657
Db 368 -----1658
QY 1658 TTTCCAACACATTATAAGATTTGTGGGATAACAAGCATGTTTAAATGCTTGTTCAGAA 1717
```





Db 673 LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys 692  
 QY 2867 ACCTTTGGGACTCAGAAATGTATGAGAGACATTTACAGCATGCACATTTCTCTTACTCAA 2926  
 Db 693 SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln 712  
 QY 2927 AGAAACTCACTGTGGA 2944  
 Db 713 LysAspValThrLeuGly 718  
 RESULT 15  
 US-09-511-507-10  
 ; Sequence 10, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 10  
 ; LENGTH: 726  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE:  
 US-09-511-507-10

Alignment Scores:  
 Pred. No.: 2,73e-38 Length: 726  
 Score: 463.00 Matches: 207  
 Percent Similarity: 38.11% Conservative: 123  
 Best Local Similarity: 23.90% Mismatches: 278  
 Query Match: 6.42% Indels: 258  
 DB: 4 Gaps: 37

US-09-302-812-3 (1-4069) x US-09-511-507-10 (1-726)

QY 599 TATTTGAACCATCAGACTCAGCAATGTCTAAGTGGCAAAATGAAGGGAACACACACG 658  
 Db 27 PheAlaHisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrThr 46  
 QY 659 GAGCAGCTTTTGGAAAGTGAACCTCAACAGTAACCCCTGGTACACAGAGAGTTTGTAGTAAT 718  
 Db 47 GluSerLysGluAsp-----ProGluGluProLysSer 57  
 QY 719 GCTAATCATGTCGCTACCTCAAAATGATGATCATGATGATGATGATGATGATGATGATGAT 778  
 Db 58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77  
 QY 779 -----AGACACATCAACAGTTTCTCACA---ACTGTAAAGCTT 814  
 Db 78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97  
 QY 815 GCAATGCAAAAGCAG-----ACTCGGAAGATGAACACGCCAGAGAACCCAAA 862  
 Db 98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp----- 115  
 QY 863 AGCCACCAAGTGCAGCAAGCTTGTCCATCTCTGGGGAAGATGTGCAAGTTGTGACGAA 922  
 Db 116 AsnHis-----LysSerThrGluProMetGluGlu----- 125  
 QY 923 GATGAGATAGACGTGTGCTCAAGAGAGTCCATTTGATGATGATGATGATGATGATGATGAT 982  
 Db 126 -----AspValAsnAsnLysSerAsnIle-----AspValAla 136

QY 983 ACTGGTGCAAAAATGATGACACAAATTTGATTATGACACAGAAAGTTCCTAGGAAATTCCTCT 1042  
 Db 137 IleAsnSerAspGluAspAspGluLeuValLeuGluGluAsn----- 150  
 QY 1043 CCATTTGAGAGAGAAAGTGAACCCGATCCACCGATGATGATGATGATGATGATGATGATGAT 1102  
 Db 151 -----AsnLysGluMetArg-----AspGlyGluGlnValGlnGlnLeu 163  
 QY 1103 TGTCAAGACTCAGAAGCAGATGAGGAG-----ACAAGTCCAGGTTTGTGATGAACAA 1153  
 Db 164 SerGlnAspLeuPheAlaAspAspGlnGluLeuLeuGluTyrProGlyIleMetLysAsp 183  
 QY 1154 GAA-----GATGGTAGTTCCTCCCAACACAGCAATAAATACCTTCACAGG 1195  
 Db 184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203  
 QY 1196 TTCCAAGCAAGAGACGCTGACATTTGAAATTT-----AGAAACCGG 1234  
 Db 204 IleGluGlnThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223  
 QY 1235 TACTCTACTAAGGC----- 1249  
 Db 224 ValArgThrSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243  
 QY 1250 GGTCAAGTTAGATTACAT-----TTCCAATTTGAAGGAGA 1285  
 Db 244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263  
 QY 1286 GAGATGCGACTGGAATGATGATTTAAATCTAAACTACCTCGA----- 1330  
 Db 264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282  
 QY 1331 -----AATATTTCTAGCTGAATGTAGATGATGATGATGATGATGATGATGATGATGAT 1378  
 Db 283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302  
 QY 1379 AAAAAGGATTTCAAAATCAGATCATTTGATGATGATGATGATGATGATGATGATGATGAT 1438  
 Db 303 ProLys-----IleValLeuPro-----GlnArg 310  
 QY 1439 AAAGAACAAGTGGGAAACCAACATCAAAAGACAGAA-----AGGAAG 1480  
 Db 311 TrpArgGluPheAspSerArgGlyArgArgArgAspSerTyrPheTyrPheLysArgLys 330  
 QY 1481 ATCCCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAGTGGCTTGGAACTCCCAT 1540  
 Db 331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350  
 QY 1541 GAGAGATGAGAAGAATG-----CCTCGGTGGGATCCGGCTGCTCTCTTCGAGACCATCT 1597  
 Db 351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu----- 368  
 QY 1598 GCCAATCACACAGTAACATTTTCGGGTAGATCTTTTGGGAGCAGGAGAGTTTCTTAAACCT 1657  
 Db 368 ----- 368  
 QY 1658 TTCCAACACATTAAGAATTTGTGGGATGAACAGCATGTTAAATGCCCTTTGTTTCAGAA 1717  
 Db 369 -----MetTyrTyrLysGluMet----- 374  
 QY 1718 CAAATTTGTACCCAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1777  
 Db 375 -----SerGlu 376  
 QY 1778 CTCATT---CAGACTGCACCTTCTCAACAAATTTACAGACCCCAAACTTGAAGATGCT 1834  
 Db 377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla----- 394  
 QY 1835 ATCTGAAATACATGTGGCATATTTCTAAGAAATGGGACCTTACAGCTTTCATCGATTC 1894  
 Db 394 ----- 394  
 QY 1895 TGGGTAAGGTACTTTGAAGACAGCAAGCTCAACATTTATATATATATATATATATATATATAT 1954

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model  
Run on: May 26, 2004, 17:10:29 ; Search time 169.187 Seconds  
(without alignments)  
13416.302 Million cell updates/sec

Title: US-09-302-812-3  
Perfect score: 7208  
Sequence: 1 ggcgtctggaagtggaggg.....agaaaaaaaaaaaaaa 4069

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1149313 segs, 278921704 residues  
Total number of hits satisfying chosen parameters: 2298626

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool.p/6333148/runat.26052004.150056.5800/app.query.fasta\_1.12437  
-DB=Published Applications AA -QPMT=fastan -SUFFIX=n2p\_rapb -MINMATCH=0.1  
-LOOPEXT=0 -UNIT5=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=6333148 @CGN 1.1.79 @runat.26052004.150056.5800  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/ECT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/ECTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	5190	72.0	976	9	US-09-973-451-4	Sequence 4, Appli
2	4635.5	64.3	977	9	US-09-973-451-2	Sequence 2, Appli
3	4395	61.0	968	9	US-09-973-451-6	Sequence 6, Appli
4	1081	15.0	768	9	US-09-973-451-8	Sequence 8, Appli
5	595	8.3	546	12	US-10-425-114-60000	Sequence 60000, A
6	463	6.4	726	9	US-09-973-451-10	Sequence 10, Appli
7	217.5	3.0	200	12	US-10-424-599-184988	Sequence 184988,
8	188	2.6	180	12	US-10-424-599-156445	Sequence 156445,
9	186.5	2.6	2476	10	US-09-824-574-7	Sequence 7, Appli
10	179.5	2.5	300	12	US-10-424-599-233915	Sequence 233915,
11	177	2.5	1163	12	US-10-221-278-197	Sequence 197, App
12	177	2.5	1163	15	US-10-291-172-197	Sequence 197, App
13	175	2.4	1111	9	US-09-815-242-12955	Sequence 12955, A
14	174	2.4	1274	12	US-10-282-122A-43837	Sequence 43837, A
15	173.5	2.4	748	9	US-09-864-761-43244	Sequence 43244, A
16	173	2.4	1948	14	US-10-032-585-7611	Sequence 7611, Ap
17	172.5	2.4	498	12	US-10-424-599-196154	Sequence 196154,
18	167	2.3	31	9	US-09-973-451-11	Sequence 11, Appli
19	166	2.3	1163	10	US-09-932-257A-3	Sequence 3, Appli
20	160	2.2	1164	15	US-10-369-493-6564	Sequence 6564, Ap
21	160	2.2	1957	15	US-10-369-493-2070	Sequence 2070, Ap
22	159	2.2	1462	14	US-10-287-218-17	Sequence 17, Appli
23	158.5	2.2	6642	15	US-10-369-493-5013	Sequence 5013, Ap
24	158	2.2	1297	12	US-10-282-122A-71692	Sequence 71692, A
25	158	2.2	1974	15	US-10-369-493-6395	Sequence 6395, Ap
26	156	2.2	1501	9	US-09-924-154-17	Sequence 17, Appli
27	156	2.2	2515	12	US-10-276-774-2651	Sequence 2651, A
28	152	2.1	1182	12	US-10-282-122A-53445	Sequence 53445, A
29	152	2.1	5560	12	US-10-263-929-142	Sequence 142, App
30	151.5	2.1	903	12	US-10-282-122A-52328	Sequence 52328, A
31	151.5	2.1	1401	16	US-10-287-226-142	Sequence 224071,
32	150.5	2.1	408	12	US-10-424-599-224071	Sequence 224071,
33	150.5	2.1	1169	12	US-10-282-122A-71179	Sequence 71179, A
34	150.5	2.1	1312	15	US-10-369-493-1950	Sequence 1950, Ap
35	150	2.1	407	12	US-10-282-122A-71177	Sequence 71177, A
36	150	2.1	1905	15	US-10-259-194A-86	Sequence 86, Appli
37	149.5	2.1	691	15	US-10-108-260A-4345	Sequence 4345, Ap
38	149.5	2.1	1210	14	US-10-205-823-664	Sequence 264, App
39	149.5	2.1	1232	14	US-10-116-712-664	Sequence 664, App
40	149.5	2.1	1232	14	US-10-116-712-669	Sequence 669, App
41	149	2.1	26	9	US-09-973-451-20	Sequence 20, Appli
42	149	2.1	29	9	US-09-973-451-12	Sequence 12, Appli
43	148.5	2.1	665	10	US-09-820-843A-107	Sequence 107, App
44	148.5	2.1	677	14	US-10-060-036-4552	Sequence 4552, Ap
45	148.5	2.1	677	15	US-10-295-027-1306	Sequence 1306, Ap

ALIGNMENTS

RESULT 1  
US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT

```

; ORGANISM: Homo sapiens
; FEATURE:
US-09-973-451-4

Alignment Scores:
Pred. No.: 0 Length: 976
Score: 5190.00 Matches: 976
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 72.00% Indels: 0
DB: 9 Gaps: 0

US-09-302-812-3 (1-4069) x US-09-973-451-4 (1-976)
QY 167 ATGAATCGGGCCCCGGCTGTGAACCTTGACCTGACCAAGCGAGCCCGCTGGGGCGCGTACA 226
Db 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20
QY 227 ACTTCGGCGGCTGCTTCGGACGCCCGGAGCTTTCGGAGCGAGCGAGCGCGTCTCGAC 286
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp 40
QY 287 CCCAAGACCGCTCAGTCGAGTTCAGGTCCACCGTCCCTCGCAGCGCTGCTCCAGGG 346
Db 41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60
QY 347 CAGCGGGACAGCAGAGGCGAGCGCCACCTCGCTTGTTCACCAAAAGACTATTACC 406
Db 61 GlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80
QY 407 AGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466
Db 81 SerTrpMetAspThrLysGlyLysThrAlaGluSerGluSerLeuAspSerLysGlu 100
QY 467 AACACAATACAGAATAGAAATCCATGATGATGTTCTGTACAAAAGATACTTTACAA 526
Db 101 AsnAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTy-Gln 120
QY 527 CATATGTAGAAAAATTAGTAAAAATTCTCAGCTAAGTCTTGTATAGTCACTCACTGAA 586
Db 121 HisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThrGlu 140
QY 587 AAAAGTACACAGTATTGAAACAGCATCGAGCATCGCAATGCTGTAAGTGGCAAAATGAA 646
Db 141 LysSerThrGlnTyLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsnGlu 160
QY 647 GGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGATAACCTGGTACCAAG 706
Db 161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180
QY 707 CAGTTTAGTAATGCTAACATTGATCGGTACCTCAAAATGATGATCACAGTGACACAGAT 766
Db 181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200
QY 767 AGTGAAGAGATAGAGACAATCAACAGATTCTCAACACTGTAAAGCTTGCAAAATGCCAAAG 826
Db 201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAsnAlaLys 220
QY 827 CAGACTACGGAAGTGAACACGCCAGAGAACCAAGCCACCGAGAGTGCAGCAGTCT 886
Db 221 GlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240
QY 887 TGCATCTCTGGGAGACTGTGCAAGTTGTGCAAGATGAGATAGACGTGCTGCCAAAG 946
Db 241 CysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProLys 260
QY 947 AGTCCATTCTCAGATGTTGGCTCTCAGGATGTTGGTACTGGGTCAAAAATCAACAACA 1006
Db 261 SerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsnLys 280
QY 1007 TTGATTAGACAAGAAGTTGCTAGAAATCTCCTCCATTGAGAGAAAGTGAACCC 1066
Db 281 LeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGluPro 300

```

```

QY 1067 GAATCACCGATGATCTGGTAAATTTCTAAAAATAGTTGTCAAGACTCAGAGCAGATGAG 1126
Db 301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320
QY 1127 GAGACAAAGTCCAGGTTTTGATGAACAAGAGATGGTAGTTCTCTCCCAAAACAGCAAAATAA 1186
Db 321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsnLys 340
QY 1187 CTTTCAAGTTCCAAAGCAGACGCTGACATTGAATTTAGGAAACCGTACTCTACTAAG 1246
Db 341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTySerThrLys 360
QY 1247 GCGGGTCAAGCTAGATTACATTTCCATTTGAGGAGGAGAGATCCACTCGAATCAAT 1306
Db 361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380
QY 1307 GATTTAAATCTCTAACTTACCTCGAAATATTTCTAGCCCTGAATGTAGAATGCAGAAATCT 1366
Db 381 AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer 400
QY 1367 AAGCAACATGGAAAAAGAGATTCTTAATACACAGATCATTTGATGAGACTGCCCAAGCA 1426
Db 401 LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla 420
QY 1427 GAGCACAGAAAGAAAGAAACAGTCGGAACCAACATCAAAAGAACAGAAAGAGATCCCT 1486
Db 421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440
QY 1487 AAATACGTTCCACTCACTTTCTCCAGATAAGAAAGTGGCTGGAACTCCCATTTGAGAG 1546
Db 441 LysTyValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu 460
QY 1547 ATGAGAAGATGCTCGTCTGGATCCGCTCTCTTGAGACCATCTGCCAATCAC 1606
Db 461 MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis 480
QY 1607 ACAGTAATCTCGGTAGATCTTTTCGAGCAGGAGAGAGTTCCTAAACCTTTTCCAACA 1666
Db 481 ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr 500
QY 1667 CATTTAAAGATTGTGGATAACAAGCATGTGTTAAATGCTTGTTCAGAACAAAAATTG 1726
Db 501 HisTyLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu 520
QY 1727 TACCCAGTGAAGATGAGATGTCGCGAGCTCGGGGAGCGGTCGAGCTCATTCAG 1786
Db 521 TyProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln 540
QY 1787 ACTGCATCTCTCAACAAATTTACACGCCCAACCTTGAAGGATGCTATTCTGAATAC 1846
Db 541 ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTy 560
QY 1847 AATGTGCAATTTCTAAGAAATGGGACTTTTACAGCTTTGATCGATTCTCGGATAAGGA 1906
Db 561 AsnValAlaTySerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal 580
QY 1907 CTTGAACAAGCAGAGCTCAACATTTATATCAGTCCATCTTGTCTGATATGGTGCAAAAT 1966
Db 581 LeuGluGluAlaGluAlaGlnHisLeuTyGlnSerIleLeuProAspMetValLysIle 600
QY 1967 GCATCTGTCTGCCAAATATTTGACCCAGCCCAATACCTCTCTGAAACAGAGATGAAT 2026
Db 601 AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn 620
QY 2027 CATTCATCACAATGTCGAGGACAGATGCGCAGTCTTTTAGCTAATGCTTCTCTGCG 2086
Db 621 HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys 640
QY 2087 ACATTTCCAGCAGAAATGCTAAGATCAATCGAGATTTCTAGTACCAGACATTAAC 2146
Db 641 ThrPheProArgAsnAlaLysMetLysSerGluTySerSerTyProAspIleAsn 660

```



```
QY 2147 TTCAATCGATTGTTGAGGACGCTTCATCAAGGAAACCGGAGAACTTAAACGCTCTTC 2206
Db 661 PheAsnArgLeuPheGluGlyArgSerArgGlySerProGluLysLeuLysThrLeuPhe 680
QY 2207 TGTACTTTAGAGAGTCACAGAGAAAAACCTACTGGGTGGTGACATTTACAGACAG 2266
Db 681 CysTyrPheArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700
QY 2267 AGTCTTGAAGATTTCCAGAAATGGAAAGATGTGAAAAACCCCTGCACCATTTGCATGTC 2326
Db 701 SerLeuGluAspPheProGluThrGluArgCysGluLysProLeuThrArgLeuHisVal 720
QY 2327 ACTTACAGAGTACCATAGAGAAATGGCCAAAGCATGCTACAGTGGATTTTGCAAAT 2386
Db 721 ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn 740
QY 2387 CGTTTGTGGAGGTGGTGAACAGTGCAGGACTGTGCAAGAAAGAAATCCGCTTTTAA 2446
Db 741 ArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPheLeu 760
QY 2447 ATCAATCTGAGTTGATTATTTTACCGCTCTTCACTGAGTGTGTCATCAATGATGT 2506
Db 761 IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780
QY 2507 CTAATTTATCACAGTACTGACGAGTACAGTGAATACACAGGCTATCTGAGACATATCGT 2566
Db 781 LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800
QY 2567 TGTCTCCGAGCACCAGAAAGTGGAGTGAAGAGGAGCTGCGAGCGGCGCTCACTGAG 2626
Db 801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820
QY 2627 ATCGTTCGATCATGCTCTTCACTTACAGCGCTACTCGATCAGTCTTGTGCTGAGAAA 2686
Db 821 IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys 840
QY 2687 ATGAGCGCGAGCTGAACAGCTTACTGTGTGATTCTCGCTCGCTGAGATTTCTTCAGAG 2746
Db 841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860
QY 2747 AATCTTTCGAGTGCACAGAACTGGGCTGGTGGCTTGGGGGTGATGCCAGG 2806
Db 861 AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyAspAlaArg 880
QY 2807 TTAAGAGCTTAATACAGATATTGGCAGCTGCTGAGCTGAGCGAGATGTGCTTTATTTC 2866
Db 881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyrPhe 900
QY 2867 ACCTTTGGGACTCAGATTGATGAGACATTTACAGCATGCAATTTCTCTTACTGAA 2926
Db 901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu 920
QY 2927 AGGAAACTCAGTTGGAGATGTATAGCTGTGCTAGCTGTGCTACGATACATCAAGCATGC 2986
Db 921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluCys 940
QY 2987 AGAACTGTTCACCCCTGGACCAACATCAAGCTTTATCCATTATACATACATCGTCTGC 3046
Db 941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960
QY 3047 GAGTCTGTGAGACGCGCTGACCATTCAGGCGCAAGGACGAGCC 3094
Db 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976
```

## RESULT 2

```
US-09-973-451-2
; Sequence 2, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
```

```
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES.
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-973-451-2
Alignment Scores:
Pred. No.: 0 Length: 977
Score: 4635.50 Matches: 874
Percent Similarity: 92.83% Conservative: 32
Best Local Similarity: 89.55% Mismatches: 69
Query Match: 64.31% Indels: 1
DB: 9 Gaps: 1
```

US-09-302-812-3 (1-4069) x US-09-973-451-2 (1-977)

```
QY 167 ATGAATGGGGGCCCCGGCTGTGAACCCCTGCACAAAGCGACCCGCTGGGGCGCGCTACA 226
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20
QY 227 ACTTCG---CCGGCTGCTTCGACGCCCGGAGCTTCCGAGCAGCAGAGGGCGCTCCTC 283
Db 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40
QY 284 GACCCCAAGGAGCTCACGTGCAGTTCAGGGTCCACCGCTCTCGCAGCCCTGCGTCCCA 343
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60
QY 344 GGGCAGGGCGGACAGCACAGAGCGCGCCACCTCGCTGTTTTCACAAACAAAGACTATT 403
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80
QY 404 ACCAGTGTGATGGACACTAAAGGAATCAAGACAGCGGATCAGAAAGTTGGATAGTAAA 463
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100
QY 464 GAAACAAACAATACAGAATAGAAATCCATGATGAGTCTGTACAAAAAGATAACTTTTAC 523
Db 101 GluAsnAsnThrArgGluSerMetSerSerValGlnLysAspAsnPheTyr 120
QY 524 CAACATAATGTAGAAAAATTAGTAATCTTCTAGCTTAAGTCTTGTGATAAGTCATCTACT 583
Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140
QY 584 GAAAAAGTACACAGTATTTTGAACCCAGCATCAGACTCGAGCAATGTGTAAAGTGGCAAAAT 643
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160
QY 644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTAGAACCTCAACAGTAGTAACCTGTGTACCA 703
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180
QY 704 GAGCAGTTTAGTAATGCTAACTGATTCGCTCAGCTCAAAATGATGATCAGATGACACA 763
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200
QY 764 GATAGTGAAGAGAATAGAGACAATCAACAGTTTCTCACAACCTGTAAAGCTTGGCAAAATCCA 823
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
```



Db 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
QY 3044 GTCGAGTCCCTGTCAGAGCCGCTGCACATTCAGGGCAAGGACAGG 3091  
Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

## RESULT 3

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973, 451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302, 812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083, 768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

## Alignment Scores:

Pred. No.: 0 Length: 968  
Score: 4395.00 Matches: 837  
Percent Similarity: 90.99% Conservative: 52  
Best Local Similarity: 85.67% Mismatches: 78  
Query Match: 60.97% Indels: 10  
DB: 9 Gaps: 6

US-09-302-812-3 (1-4069) x US-09-973-451-6 (1-968)

QY 167 ATGAATCGCGGCCCGCTGTGAACCTGACCAAGCAGCCGCTGGGCGCGCT--- 223  
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAla---ArgTrpGlyAlaAlaGly 19  
QY 224 ACAACTCGCGCGCTGCTTCGGACGCCCGGAGCTTTCCGAGCAGGCGCGCTCCTC 283  
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
QY 284 GACCCAGGCGCTCAGCTGCAAGTTCAGGTCGCCACCTCGCTCGCAGCGCTCGTCCCA 343  
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59  
QY 344 GGCAGCGCGGACAGCACAGAGGCGGCCACCTCGCTGCTTTTCAACAAAGACTATT 403  
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
QY 404 ACCAGTTGGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAAGTTTGGATAGTAA 463  
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
QY 464 GAAACACACATACAGATAGATCCATGATGAGTCTGTACAAAAGNATACCTTTAC 523  
Db 97 GluAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116  
QY 524 CAACATAATGTAGAAAATTAGTAAATGTTCTCAGCTAAGTCTGTGATAGTCACTCACT 583  
Db 117 ProHisLysValGlnLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
QY 584 GAAAAAAGTACACAGTATTTTGAACACGATCATGATGCGAGCAATGTGTAAAGTGGCAAAAT 643

Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156  
QY 644 GAAGGAAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAAACAGTAAACCTGGTACCA 703  
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
QY 704 GAGCAGTTTAGTAATGCTAACTTGCATTCGCTACCTCAAAATGATCATCAGTGACACA 763  
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
QY 764 GATAGTGAAGAGATAGACACATCAACAGATTTCTCACAACCTGTAAGCTTGCATGCA 823  
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProLysLysLeuAlaAsnThr 216  
QY 824 AAGCAGACTACGGAAGATGAACACGCCAGAGAAGCCAAAGCCACCAAGAGTGCAGCAAG 883  
Db 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233  
QY 884 TCTTGCATCCTCGGGAAGACTGTGCAAGTTGTGCACAAGATCAGATAGACGTGGTGCCA 943  
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGlnGluValAspValLeuPro 253  
QY 944 AAGAGTCCATTTGTGAGATGTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAC 1003  
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
QY 1004 AAATTGATTAGACAAAAGTTGCCCTAGGAATTTCTCTCCATTTGAGAAGGAAGTGA 1063  
Db 274 LysLeuThrGlyGlnGluSerLeuGlyAspSerProProPheGlnLysGluSerGlu 293  
QY 1064 CCCGAATCACCGATGATGATGATTAATCTTAAATAATAGTTGTCAAGACTCAGAAAGCAGAT 1123  
Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
QY 1124 GAGGAGACAAGTCCAGGTTTGTGATGAACAAGAAAGATGGTAGTTCTCCCAACAGCAAAAT 1183  
Db 314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332  
QY 1184 AAACCTTCAAGGTTCCAGCAGAGACGCTGACATTTGAATTTAGGAAACGGTACTCTACT 1243  
Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
QY 1244 AAGGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGAGGAGAGAGTCCGACTGGAATG 1303  
Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371  
QY 1304 AATGATTTAAATGCTAAACTACTCGGAATATTTCTAGCTGAAATGTAGATGCAGAAAT 1363  
Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
QY 1364 TCTAGCAACATGGAAAAAGGATCTTAAATACACAGATCATTTGATGAGACTGCCCAAA 1423  
Db 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411  
QY 1424 GCAGAGGACAGAAAGAAAACAGTGGGAAACCAACATCAAGAAACAGAAAGGAAGATC 1483  
Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
QY 1484 CCTAATAGCTTCCACTCACCTTTCTCCAGATAGAAAGTGGCTTGGAACTCCCATGAG 1543  
Db 432 ProLysTyrIleProProAsnLeuProProGluLysLysIleTrpLeuGlyThrProIleGlu 451  
QY 1544 GAGATGAGAAGAAATGCTCGGTGGGATCCGCTGCTCTCTTGAGACATCTGCAAT 1603  
Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
QY 1604 CACACAGTAACTATTGGGTAGATCTTTTTCGAGCAGGAGAAAGTTCCTTAAACCTTTTCA 1663  
Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
QY 1664 ACACATTATAAGATTGCGGATTAACAGCATGTTAAATGCTTGTTCAGACAAAT 1723

```

Db      492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511
QY      1724 TTGTACCCAGTGGAGATGAGAACTGTCAGGAACTGCGGGAGCGGTGGAGCTCAT 1783
Db      512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuile 531
QY      1784 CAGACTGCACCTCTCAACAAATTTACAGACCCCAAACTTGAAGGATGCTATTCTGAA 1843
Db      532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551
QY      1844 TACAATGTGGCATATTTCAAGAAATGGGACTTTACAGCTTTGATCGATTTCTGGGATAAG 1903
Db      552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571
QY      1904 GTACTTGAAGACGACAGAGCTCAACATTTATATATATAGTCCATCTTGCCTGATATGCGTAA 1963
Db      572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591
QY      1964 ATTGCACCTCTGTCGCCAAATATTTGCCACCCAGCCCAATACCACTCTCTGAAACAGAGATG 2023
Db      592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611
QY      2024 AATCATTTCCATCACAAATGTCGACGAAACAGATGTCAGTCTTTTAGCTAATGCTTTCTTC 2083
Db      612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631
QY      2084 TGCATATTTCCACGACGAAATGCTAAGATGAATCGGAGTATTCTAGTTACCCAGACATT 2143
Db      632 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651
QY      2144 AACTTCATTCGATGTTTGGGAGCTTCATCAGAGAAACCGGAGAACTTAAACGCTC 2203
Db      652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671
QY      2204 TTCTGCTACTTTAGAGAGTCAACAGAGAAACCTACTGCTGTTGTCACATTTACAGA 2263
Db      672 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691
QY      2264 CAGAGCTCTGAAGATTTTCCAGAAATGGGAAAGATGTGAAACCCCTTGACAGATTGCAT 2323
Db      692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711
QY      2324 GTCACTTACGAGTACCATAGAGAAATGCCAAGCATGCTACAGGTGGATTGCA 2383
Db      712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
QY      2384 AATCGTTTTGTTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2443
Db      732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751
QY      2444 TTAATCAATCTGAGTTGATTATTTACGGCTCTTTCAGTGGTGGATGATCAATGAA 2503
Db      752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
QY      2504 TGCTTAATATACAGGTACTGACGATGACGATGACATGATACAGGCTATGCTGACACATAT 2563
Db      772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791
QY      2564 CGTTGTCCCGGACCCAGAGATGGGAGTGAAGAGGACGACTCGGAGCGCGCTGCAC 2623
Db      792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgArgCysThr 811
QY      2624 GAGATCGTTGCATCGATGCTCTTCATTTACAGCTTACCTCGATCGATTTGTGCGCTGAG 2683
Db      812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831
QY      2684 AAAATGAGACGCGAGCTGAACAGCTTACTGTGATTTCTCCGCTCGTGGATTTCTTCA 2743
Db      832 LysValArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851
QY      2744 GAGAAATCTTTTCAGTGGCCACAGGAACTGGGCGCTGTGCTGCTTTGGGGGTGATGCC 2803
Db      852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

```

```

QY      2804 AGTTTAAAGCTTTAATACAGATATTGGCAGCTGCTGAGCTGAGCGAGATGTGTTTAT 2863
Db      872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891
QY      2864 TTACCTTTGGGAGCTCAGAAATTCATGAGAGACATTTACAGCATGCACATTTTCTTACT 2923
Db      892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
QY      2924 GAAAGGAAACTCACTGTTGGAGATGTGTATAGCTGTTGTCTACCATCTACAAATGAAGA 2983
Db      912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931
QY      2984 TGCAGAACTGTTCCACCCCTGGACACATCAAGCTTTTATCCATCATATACCATGCT 3043
Db      932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951
QY      3044 GTCCAGTCTCTGTGCAGACCGCTGACCATTCAGGGGAAAGGACAGGGACC 3094
Db      952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 4
; Sequence 8, Application US/09973451
; Patent No. US2002013238A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-973-451-8

Alignment Scores:
Pred. No.: 1,92e-77 Length: 768
Score: 1081.00 Matches: 246
Percent Similarity: 57.89% Conservative: 99
Best Local Similarity: 41.28% Mismatches: 197
Query Match: 15.00% Indels: 54
DB: 9 Gaps: 16

US-09-302-812-3 (1-4069) x US-09-973-451-8 (1-768)
QY      1412 AGACTGCCCAAGACAGGACAGAGAAAGACAGTGGGAAACCAACATCAAGAACA 1471
Db      45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu----- 62
QY      1472 GAAAGGAAGATCCCTAAATACATGTTCCACCTCACCTT-----TCTCCAGATAAGAAGTGG 1525
Db      63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTip 74
QY      1526 CTTGGAACTCCCATTTGAGGAGATG-----AGAAGAATGCTCGGTGGGATCCGGCTG 1579
Db      75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94
QY      1580 CCTCTCTTGAGACCATCTGCCAATCATCACAGTAATATTTCGGGTAGATCTTTTTCGAGCA 1639

```

Db 95 ProProValThrAlaGlyAsnLeuLeuHisArgValMetTyrGlnLeuProIleArgGluThr 114  
 QY 1640 GGAGAAATCTCTAAACCTTTTCCACACATATATAAGATTGTGGGATACACAGCATGTT 1699  
 Db 115 -----ProProArgProTyrLysSerProGlyLys-----TrpAspSerGluHisVal 130  
 QY 1700 AAATGCTTTGTTTCAGAACAAATTTTACCCAGTGAAGATGAGATGCTGAGCGAACT 1759  
 Db 131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150  
 QY 1760 GCGGGAGCGGTGGAGCTCATTCAGACGTGCATCTTCAACAAATTTACACAGCCCAA 1819  
 Db 151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu 170  
 QY 1820 AACTGAGAGATGCTATTCTGAATAACAATGTGGCATATTCTAAGAAATGGACTTTACA 1879  
 Db 171 GluLeuGlnAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190  
 QY 1880 GCTTTGATCGATTCTCGGATAGGTACTTGAAGAAGCAGAGCTCAACATTTATATCAG 1939  
 Db 191 AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGlnThrArgValPhePheGlu 210  
 QY 1940 TCCATCTTGCTGATATGTTGAAATATGCACTCTGCTGCCAATATTTGCCACCCGCCA 1999  
 Db 211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230  
 QY 2000 ATACCCTCTGAAACAGAGATGAATCATTCATCATCAATGTCGAGGAAACAGATGCC 2059  
 Db 231 ValProLeuLeuLysHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250  
 QY 2060 AGTCTTTAGCTAATGCTTTCTCTGACATTTCCACGACGAAATGCT---AAGATGAAA 2116  
 Db 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270  
 QY 2117 TCGAGATTTCTAGTACCCAGACATTAACCTCAATCGATTGTTTGAGGACGCTTCATCA 2176  
 Db 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290  
 QY 2177 AGGAACCGGAGAACTAAACGCTCTCTGCTACTTTAGAGAGTC-----ACGAG 2230  
 Db 291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310  
 QY 2231 AAAAAA-----CCTACTGGGTGGTGCACATTTTCAAGACAGAGT-----CTT 2272  
 Db 311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgSerGlyLeuPro 330  
 QY 2273 GAAGATTTTCCAGATGGGAAAGATGGAACCCCTTG-----ACAGATTGCATGTC 2326  
 Db 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350  
 QY 2327 ACTTAGAGGTACCATAGAGAAATGGCCAGGCATGTCTACAGTGGATTTTGCAT 2386  
 Db 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370  
 QY 2387 CGTTTGTGTGGAGTGGTGTAAACAGTCAGGACTGTGCAAGAAGAAATCCGCTTTTA 2446  
 Db 371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal 390  
 QY 2447 ATCAATCTGAGTTGATTATTTCACGGCTCTTCTACAGTGGTGTGATCAATGAATGT 2506  
 Db 391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAla 410  
 QY 2507 CTAATATACAGGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566  
 Db 411 LeuValMetLeuGlyAlaGluArgGlySerAsnTyrThrGlyTyrAlaGlySerPheGlu 430  
 QY 2567 TGCTCCGGAGCCACGAGATGGAGTGAAGGACGACTGCGAGCGCGCTCATCTGAG 2626  
 Db 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450  
 QY 2627 ATCGTTGCCATCGATGCTCTTCTACTTCAGACGCTACCTCGATCGATGTTGGCTGAGAA 2686  
 Db 451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeu 470

QY 2687 ATGAGCGCGAGCTGACACAGGCTTACTGTGGATTTCTCCGT----- 2728  
 Db 471 MetGluArgGluLeuAsnLysAlaIleGlyPheValHisTrpMetValThrProPro 490  
 QY 2729 CTTGGAGATTTCTTTCAGAGATCTTCTGCGATGCGCCACAGAACTGGGCTGTGGTGGC 2788  
 Db 491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502  
 QY 2789 TTTGGGGGTGATGCCAGGTTAAAGCTTAATACAGATATTGGCAGCTGTCGACGCTGAG 2848  
 Db 503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522  
 QY 2849 CGAGATGTGTTTATTTACCTTTGGGACTCAGAATTGATGAGACATTTACAGCATG 2908  
 Db 523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542  
 QY 2909 CACATTTCTTACTGAAAGAACTCCTGTTGGAGATGTGTATAGCTGTGTGTGTACGA 2968  
 Db 543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561  
 QY 2969 TACTCAATGAAGAATGACAGAACTGTTCCACCCCTGGACCA----- 3010  
 Db 562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581  
 QY 3011 GACATCAAGCTTTATCCATTCATATACCATGCTGTCTGAGTCTGTGCGAGACCGCTGAC 3070  
 Db 582 LysLysLysLeuTyrAspPheIle-----LysGluGluLeuLysValAlaArg 598  
 QY 3071 CATTCAGGGCAA---AGGACAGGACCTGAGGAGCGCAGCGCATAGCA 3115  
 Db 599 ValProGlyGluClyAlaSerAlaGluAlaGlySerSerArgValAla 614

## RESULT 5

US-10-425-114-60000  
 ; Sequence 60000, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jindong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovacic, David K.  
 ; APPLICANT: Screen, Steven E.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 60000  
 ; LENGTH: 546  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3689-262-E12\_FLI.pep  
 US-10-425-114-60000

Alignment Scores:  
 Pred. No.: 1,5e-38 Length: 546  
 Score: 595.00 Matches: 151  
 Percent Similarity: 50.34% Conservative: 72  
 Best Local Similarity: 34.09% Mismatches: 150  
 Query Match: 8.25% Indels: 70  
 DB: 12 Gaps: 12

US-09-302-812-3 (1-4069) x US-10-425-114-60000 (1-546)

QY 1823 TTGAAGGATGCTATTCTGAAATACATATGCGCATATTTCT-----AAG 1864  
 Db 88 LeuAlaAspAlaLeuThrAspLeuArgLeuAlaLeuProAlaLeuProAlaLeuProArg 107  
 QY 1865 AAATGGGACTTTACAGCTTTGATCGATTTCTGGGTAAGGTACTTGAAGAGCAGAGCT 1924

```
Db      108 AlaAlaGspGlyLeuAlaLeu-----PhePheAspLeuLeuSerArgAlaGlnAla 125
Qy      1925 CAACATTTATATACAGTCCTGCTGATATGCTGAAATGCACTCTGCTGCAAT 1984
Db      126 ArgGlyTrpPheSerGluValValProAlaLeuAlaArgLeuLeuLeuArgLeuProThr 145
Qy      1985 ATTTGACCCAGCCCAATACCACTCTCTGAAACAGAGATGAATCAATTC----- 2032
Db      146 LeuLeuGluAspHisTyr-----AlaLysAlaGlyHisGlyAlaSerGlyLeu 161
Qy      2033 -----ATCACATGTCGAGAAACAGATGCGCAGT 2062
Db      162 ArgValMetAlaSerGlnAspAlaGlyValValLeuLeuLeuSerGlnGluLeuValAlaAla 181
Qy      2063 CTTTGTAGCTAATGCTTCTCTGACATTTCCAGCAGCAATGCTAAGATGAATCGGAG 2122
Db      182 LeuLeuThrCysAlaLeuPheCysLeuPheProThrAlaGlyArgAlaGln----- 198
Qy      2123 TATTCTAGTTACCCAGACATTAATCAATCGATTGTTT-----GAGGGA 2167
Db      199 ---AlaCysLeuProThrIleAsnPheAspGlyLeuPheAlaProLeuIleHisAla 217
Qy      2168 CGTTCATCAAGAAACCGAGAACTTAAACGCTCTCTGCTACTTTAGAAAGTCA 2227
Db      218 ArgGlnSerGlnGlu---GlnLysValArgCysLeuValHisTyrPheGluArgValThr 236
Qy      2228 GAGAAAAACCTACTCGGCTCGTGCATTTACAGACAGAGCTT----- 2272
Db      237 AspSerMetProThrGlyLeuValSerPheGluArgLysValLeuProArgAlaLeu 256
Qy      2273 -----GAAGATTTCAGAAATGGGAAAGATGTGMAAAACCCCTTGACA 2314
Db      257 SerAspGlyValProTyrProAspIleHisAlaTrpValAlaSerSerAlaProLeuCys 276
Qy      2315 CGATTGCATGCTACTACGAGGTACCATAGAAGAAATGCGCAAGGCATGCTACAGTG 2374
Db      277 GlnPheThrValPheSerSerGlyPheIleGluAspGluGlnGluAlaLeuGlnVal 296
Qy      2375 GATTTGCAATCGTTTGTGGAGTGGTGTACCACTGAGGCTGCTGCTGAGGTGCTGAT 2434
Db      297 AspPheAlaAsnLysTyrLeuGlyGlyAlaLeuSerTrpGlyCysValGlnGlnGlu 316
Qy      2435 ATCCGCTTTTAAATCAATCCTGAGTTGATTATTTCCAGGCTCTTCACAGGTGCTGGAT 2494
Db      317 IleArgPheMetIleAsnProGluLeuIleValGlyMetLeuPheLeuSerCysMetGlu 336
Qy      2495 CACATGATGCTTAATATACAGGTACTGACGAGTACAGTGAATACACAGGCTATGCT 2554
Db      337 AspAsnGluAlaIleGluIlePheGlyAlaGluArgPheSerGlnTyrMetGlyTyrGly 356
Qy      2555 GAGCATATCGTGTGCTCCGAGCCACGAGATGGAGTGAAGGAGGAGCTGCGAGGG 2614
Db      357 SerSerPheArgPheValGlyAspTyrLeuAspThrLysProPheAspSerMetGlyArg 376
Qy      2615 CGCTGCTAGATGCTTGGCCATCGATGCTCTTCTTACAGCGCTACCTCGATCAGTTT 2674
Db      377 ArgArgThrArgIleValAlaIleAspAlaLeuAspCysProAlaArgLeu---HisTyr 395
Qy      2675 GTGCTGAGAAATAGAGCGGAGCTGAACAGGCTTACTGTGATTTCTCGT----- 2728
Db      396 GluSerGlyCysLeuLeuArgGluValAlaAsnLysAlaPheCysGlyPheAspGlnSer 415
Qy      2729 -----CCTCGA 2734
Db      416 LysHisHisLeuTyrAlaLysLeuPheGlnAspLeuHisAsnLysAspPheSerSer 435
Qy      2735 GTTTCCTTCAGAAATCTTCTGACGTGCCACAGAAACTGGGCTGTGGTGCCTTTGGG 2794
Db      436 IleAsnSerSerGluTyrValGlyValSerThrGlyAsnTrpGlyCysGlyAlaPheGly 455
Qy      2795 GGTGATGCCAGGTTAAAGCCCTTAATACAGATATTGGCAGCTGCTGCGAGTGGAGAT 2854
```

```
Db      456 GlyAsnProGluIleLysSerMetIleGlnTrpIleAlaAlaSerGlnAlaLeuArgPro 475
Qy      2955 GTGCTT---TATTTACCTTTGGGAGCTCAGAAATGATGAGACATTTTACAGATGCAC 2911
Db      476 PheValAsnTyrTyrThrPheGluAspValSerLeuGlnArg---LeuGluGluValle 494
Qy      2912 ATTTCTCTTACTGAAGAAACTCCTGTTGGAGATGTGTATAAGCTTGTCTGCTAGATAC 2971
Db      495 GlnTrpIleArgLeuHisGlyTyrThrValGlyGluLeuTrpHisMetLeuMetGluTyr 514
Qy      2972 TACAATGAA 2980
Db      515 SerSerGln 517

RESULT 6
; Sequence 10, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: N2AD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 10
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
US-09-973-451-10

Alignment Scores:
Pred. No.: 6,41e-28 Length: 726
Score: 463.00 Matches: 207
Percent Similarity: 38.11% Conservative: 123
Best Local Similarity: 23.90% Mismatches: 278
Query Match: 6.42% Indels: 258
DB: 9 Gaps: 37

US-09-302-812-3 (1-4069) x US-09-973-451-10 (1-726)
Qy      599 TATTTGAACCAAGCATCAGACTGCGAGCAATGTGTAAAGTGGCAAAATGAAGGAAACACAC 658
Db      27 PheAlaHisGlnValProThrMetLysArgGlyLysLeuThrGluHisGlyAsnThrThr 46
Qy      659 GAGCAGCTTTTGAAGTGAACCTCAACAGTAACTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 718
Db      47 GluSerLysGluAsp-----ProGluGluProLysSer 57
Qy      719 GCTAACATGATCGGTCCCTCAAAATGATGATCAGTGCACAGTACAGATAGTGAAGAGAT 778
Db      58 ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn 77
Qy      779 -----AGGACAAATCAACAGTTTCTCACA---ACTGTAAAGCTT 814
Db      78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97
Qy      815 GCAATGCAAGACAG-----ACTACGGAAGATGAACACGCGCAGAGAGACCAAA 862
Db      98 SerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAsp----- 115
Qy      863 AGCCACCAGAGTGCAGAAAGTCTTGCCATCTCTGGGGAAGACTGTGCAAGTTGTGCAGAA 922
```





[illegible]

## RESULT 8

```

RESULT 8
US-10-424-599-156445
; Sequence 156445, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 156445
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_112291C.1.pep
US-10-424-599-156445

```

### Best Local Simila

	Length:	180
	Matches:	49
	Conservative:	23
	Mismatches:	41
	Indels:	50
	Gaps:	4
DR.		12
Alignment Scores:		
Pred. No.:	3.99e-06	
Score:	188.00	
Percent Similarity:	44.1%	
Best Local Similarity:	30.0%	
Query Match:	2.61%	

US-09-302-812-3 (1-4069) x US-10-424-599-156445 (1-180)  
QY 2696 GAGCTGAACAAGGCTTACTGTGGATTTC-----CCTCCT----- 2731

QY	2696	GAGCTGAACAAGGCTTACTGTGGATTCTC	-----GCTCCT-----	2731
Db	7	GluileasnLysalaPheCysGlyPheLeuTyrglnCysIysTyrglnPnProTyrglnLys	26	
		-----	-----	
QY	2731	-----	-----	2731

2731	-----	2731
------	-------	------

Qy	2731	-----	2732
Db	47	GlutHrAspGluGlyGluIleSerAsnHisLysIleThrAsnSerClnAsnAspTyrHis	66

Qy	2732	GGAGTTTCTTCAGAGAAATCTTTCTGCAGTGGCCACAGGAACTCTGGGCTGTGGTGCTTT	27191
----	------	--	-------

Db 67 GlyMetAspGlnGlyAsnAsnIleGlyValAlaThrGlyAsnTrpGlyCysGlyAlaPhe 86

QY 2792 GGGGGTGATCCAGGTTAAAGCCTTAATACAGATATTTGGCAGCTGCTGCAGCTGAGGA 2851

Db 87 GlyGlyAspProGluValIleThrIleLeuGlnTrpLeuAlaAlaSerGlnAlaLeuArg 106

2852 GAT---GTGGTTTATTTTACCTTTGGGACCTCAGAATTGATGAGAGACATTACAGCATG 2908

Dy	2852 GAT---
	GGTGTATTACCTTTGCGAGCAGAAATTCAGACCAATTCAGCATTAACACTA
	: : : : : : : : : : : : : : :
	GAT---
	GGTGTATTACCTTTGCGAGCAGAAATTCAGACCAATTCAGCATTAACACTA
	: : : : : : : : : : : : : : :
Dd	107 ProPheIleAlaTyrrThyrPheGly---LeuGluAlaLeuGlnSerLeuAspGluVal
	125

Db	107	ProPheIleAlaIyIyIyIThrPheGly---	LeuGluAlaLeuGlnSerLeuaspGluval	125
Qv	2909	CACATTTTCCTACTCAAAAGGAAACT	CACGTGTGGAGATGTGTATAAGCTGTGTCACGA	2968

Qy	2909	CACATTTCTTACTGAAAGGAAACTCAGCTGTTGGAGATGTATAAGCTCTTGTACGA
	126	AlaLSTTLEIeSvSvCfLaRoTThrVaGclVvAaSnLeuTAsnMetLeuLecLu
Ab		

Db 126 AlaHisTrpIleLeuSerGlnArgTrpThrValGlyAspLeuTrpAsnMetLeuIleGlu 145

QY 2969 TACTACAATGAAGAATGCAGAAACTGTTCACCCCCTGGACCAGACATCAAGCTTTATCCA 3028

Db 146 TyrSerIleAsnArgSerLysGlyGluThrAsnValGlyPheLeuGlnTrpLeuLeuPro 165





1474 euArgThrGluThrGlnAsnAlaLeuLysGluGluGluArgArgLysArgIleAlaG 1494  
 1452 AAACCAACATCAAGACAGAAAGAGATCCCTAAATACGTCCAA---CCTCACCTTT 1508  
 1494 luArgGluArgGluArg-----GluLysLeuArgGluValIleGluIleGluAAs 1512  
 1509 CTCACGATGAAGTGGCTTGGAACTCCCAT---GAGGAGATGAGAA 1553  
 1512 erProThrLysCysProIleThrThrLysLeuValLeuAspGluAsnGluThrLysG 1532  
 1554 GAATGCTCGGTGGATCCGGCTCTCTTGAGACCATCTGCCAATCAGACAGTAA 1613  
 1532 lu-----ProLeuValGlnValHisArgAsnMetValIleL 1544  
 1614 CTATTCGGGTAGATCTTTTTCGAGCAGGAGAGTCTCTAAACCTTTTCCAAACATATA 1673  
 1544 ysLeu-----LysPro-----HisGlnV 1550  
 1674 AAGAT-----TTGTGGGATAACAAGCATGTTAAATGCCCTTGTTCAGACAAA 1721  
 1550 alAspGlyValGlnPheMetTrpAsp-----CysCysCysGluS 1563  
 1722 ATTGTACCCAGTGGAGATGAGATCGTACGCACTGGCGGAGCGGTGGAGTCA 1781  
 1563 erVal-----GluLysThrLysLysSerProGlySerGlyCysIleLeuA 1578  
 1782 TTCAGACTGCACCTTCTCAACAAATTTACAGACCCCAAACTTGAAGGATGCTATTCTGA 1841  
 1578 laHisCysMetGlyLeuGlyLys-----ThrLeuGlnValValSerPheL 1593  
 1842 AATACAAATGGGCAATTTCTAAGAAATGGGACTTT---ACAGCTTGCATGATTC--- 1894  
 1593 euHisThrValLeuLeuLeuCysAspLysLeuAspPheSerThrAlaLeuValValCysProL 1613  
 1895 -----TGGATGAAGTACTTGAAGAGCAGAGCTCAACATTTATATC 1937  
 1613 euAsnThrAlaLeuAsnTrpMetAsnGluPheGluLys-----TrpG 1627  
 1938 AGTCCATCTTCTCGATATGTTGAAATTTGCACTCTGTCTGCCAAATATTTGCCACCGC 1997  
 1627 lngluGlyLeuAsnAspAsnGluLysLeuGlu-----ValSerGluLeuAlaThrValL 1645  
 1998 CAATACCACTCCTGAAACAGAAATGATCATTCATCATCAATGTCGAGCAAGACAGATTG 2057  
 1645 ysArgPro-----GlnGluArgSerTyMetLeuGlnArgTrpGlnGluAspGlyG 1662  
 2058 CCAGTCTTTAGTAAATGCTTCTCTGCATTTCCAGCAGCAATGCTAAGATGAAT 2117  
 1662 lyValMetIleIleGly----- 1667  
 2118 CGGAGTATCTAGTTACCCAGACATTAATCTCAATCGATTTGTTGAGGACGTTTCATCAA 2177  
 1668 -----TyrGluMetTy-----ArgAsnLeuAlaGlnGlyArgAsnVal- 1680  
 2178 GGAACCGGAGAACTTAAACGCTTCTCTGCTACTTTAGAGAGTACAGAGAAAC 2237  
 1681 --LysSerArgLysLeuLysAspIlePhe-----AsnLysAlaLeuValAspProGlyP 1698  
 2238 CTACTGGGTGGTGACATTTACACAGCAGAGCTTTGAAGATTTTCCAGATGGGAAAGAT 2297  
 1698 roAspPheValValCysAspGluGlyHisIleLeuLysAsn-----GluAlaSerAlav 1716  
 2298 GTGAAACCCCTTG-----ACACGATTCATGCTCACTTACGAAGTACC----- 2341  
 1716 alSerLysAlaMetAsnSerIleLysSerArgArgIleIleLeuThrGlyThrProL 1736  
 2342 -----ATAGAAGAAATGCCAAG 2360  
 1736 euGlnAsnAsnLeuIleGluTyHisCysMetValAsnPheIleLysGluAsnLeuLeuG 1756  
 2361 GCATGCTACAGTGGATTTTGCATATCGTTTGT-----GGAGGTGGTGTAA 2408  
 1756 lysIleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGlyGlnCysA 1775

QY 2409 CCACTGAGGACTTGTGCAAGAGAAATCCGCTTTTAAATCAATCCTGAGTTGATTATT 2468  
 Db 1775 laAspSerThrMetVal-----AspValArgValMetLysLysArgAlaHisIleLeuT 1793  
 QY 2469 CACGGCTC-----TTCACCTGAGGTGCTG- 2491  
 Db 1793 yrGluMetLeuAlaGlyCysValGlnArgLysAspTyThrAlaLeuThrLysPheLeuP 1813  
 QY 2492 -----GATCACAATGAATGCTTAATATATACAGGTACTGAG---CAGTACAGTGAATACA 2543  
 Db 1813 roProLysHisGluTyThrValLeuAlaValArgMetThrAlaIleGlnCysLysLeuTyG 1833  
 QY 2544 CAGGCTATCTGAGACATATCTGTCGCCGAGCCACCAAGATGGAGTGAAGGACG 2603  
 Db 1833 lnTyTyThr-LeuAspHisLeuThrGlyValGlyAsnSerThrGluGlyArgGly--- 1851  
 QY 2604 ACTGCGAGCGGCTGCACTG 2624  
 Db 1852 ---LysAlaGlyAlaLysLeu 1857  
 RESULT 10  
 US-10-424-599-233915  
 ; Sequence 233915, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovacic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)/B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 233915  
 ; LENGTH: 300  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5324C.1.pep  
 US-10-424-599-233915  
 Alignment Scores:  
 Pred. No.: 2,36e-05 Length: 300  
 Score: 179.50 Matches: 56  
 Percent Similarity: 34.24% Conservative: 45  
 Best Local Similarity: 18.98% Mismatches: 65  
 Query Match: 2.49% Indels: 129  
 DB: 12 Gaps: 5  
 US-09-302-812-3 (1-4069) x US-10-424-599-233915 (1-300)  
 QY 1892 TTCTGGGATAGTACTTGAAGAGCAGCAAGCTCAACATTATATATCATCTTCCTGCT 1951  
 Db 76 PhePheAspGluValMetSerGlyGluGluSerSerLysTrpPheGlnGluValLeuPro 95  
 QY 1952 GATATGTGAAAATGTCACCTCTGCTGCCAAATATT----- 1987  
 Db 96 AlaLeuGlyAsnLeuLeuArgLeuProSerLeuLeuGluSerHisTyrGlnAsnThr 115  
 QY 1988 -----TGCAACCCAGCAATACCACTCTCTG 2011  
 Db 116 AspAsnMetAlaIleAspGlyGluAlaGlyAlaMetLeuThrThrAlaLeuArgLeuLeu 135  
 QY 2012 AAACAGAGATGAATCATTCATCATCAATGTCGAGGAAACAGATTCGAGCTTTTACGT 2071  
 Db 136 AspSerGlnGlnProGlyIleValPheLeuThrGlnGluLeuIleAlaAlaLeuLeuSer 155  
 QY 2072 AATGCTTTCTTCGACATTTCCAGCAGCAATGCTAAGATGAATCGAGTATTCTAGT 2131  
 Db 156 CysSerLeuPheCysLeuPhePro-----ValSerAsp 166





```

QY 2747 -----AATCTTCTGCAGTGGCCACAGCAAACTGGGGCTGTGGTGCCTTT 2791
Db 1055 SerProValSerProLysLeuSerProGlyAsnSerGlyAsnIlyrSerSerGlyAlaSer 1074
QY 2792 GGGGGTGATGCC 2803
Db 1075 SerAlaSerAla 1078

RESULT 12
US-10-291-172-197
; Sequence 197, Application US/10291172
; Publication No. US20030228584A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: No. US20030228584A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-045
; CURRENT APPLICATION NUMBER: US/10/291,172
; CURRENT FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/693,267
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/665,363
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/616,847
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 09/596,193
; PRIOR FILING DATE: 2000-06-17
; PRIOR APPLICATION NUMBER: 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: 09/519,705
; PRIOR FILING DATE: 2000-03-07
; NUMBER OF SEQ ID NOS: 752
; SEQ ID NO 197
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-172-197

Alignment Scores:
Pred. No.: 6.52e-05 Length: 1163
Score: 177.00 Matches: 212
Percent Similarity: 32.78% Conservative: 144
Best Local Similarity: 19.52% Mismatches: 405
Query Match: 2.46% Indels: 325
Db: 15 Gaps: 49

US-09-302-812-3 (1-4069) x US-10-291-172-197 (1-1163)
QY 110 AGTGAATGAGGCTCTCCGGGGCGGGCCGGGACAGTGGCTGTGGT 160
Db 132 SerSerGlnArgThrSerAlaGlySerSerGlyThrAsnSerSerGlyGlnArgHis 151
QY 160 ----- 160
Db 152 AspArgGluSerTyrAsnAsnSerGlySerSerSerArgLysLysGlyGlnHisGlySer 171
QY 161 -----CCAGCATGAATCGGGCGGGCCCGGCTGTGAACCTGCACCAAGACGACCGCG 211
Db 172 GluHisSerLysSerArgSerSerProGly-----LysProGlnAla 186
QY 212 TGGGGCGCGCTACAACTTCGCGCGCTGTCTCGACGCCCGCGAGCTTTCGAGCAGGCG 271
Db 187 ValSerSerLeuAsnSerSerHisSerArgSerHisGlyAsnAspHisHisSerLysGlu 206
QY 272 ----AGCGCGTCTCTGACCCCAAGACGCTCAGTCAGTTCAGGGTCCACCGTCTCG 328
Db 207 HisGlnArgSerLysSerProArgAspProAspAlaAsnTrp-----AspSer 222
QY 329 CCAGCTGTGCTCCGGGGCGGACGACGAC----- 361
Db 223 ProSerArgValProPheSerSerGlyGlnHisSerThrGlnSerPheProSerLeu 242
QY 362 AGAGGCGGCGCACCTCGCTGTGTTTCAACCAAAAGACTATTACCATGTTGGATGACACT 421

```

```

243 MetSerLysSerAsnSerMetLeuGlnLysProThrAlaTyrValArgProMetAsp--- 261
422 AAAGGATCAAGACAGCGGAA-----TCAGAAAGTTTGGATAGTAAAGAAAC 469
262 ---GlyGlnGluSerMetGluProLysLeuSerSerGluHisTyrSerSerGlySerHis 280
470 AACAAACAAGAATAGATCCATGATGATGTTCTGTACAAAAGATAACTTTTCAACAT 529
281 GlyAsnSerMetThrGluLeuLysProSerSer-----LysAla 293
530 AATGTAGAAAATAGTAAATGTTTCTCAGCTAAGTCTTGATAAGTCA----- 577
294 HisLeuThrLysLeuLysIleProSerGln---ProLeuAspAlaSerAlaSerGlyAsp 312
578 -----CTCACTGAAAAGATACACAGTATTTTGAACACGACATCAGACT 619
313 ValSerCysValAspGluLysLeuLysGluMetThrHisSerTrpProProLeuThr 332
620 GCA-----GCAATGTGTAAAGTGGCAAAATGAA-----GGGAAACACACG 658
333 AlaIleHisThrProCysLysThrGluProSerLysPheProPheProThrLysGluSer 352
659 GAGCAGCTTTTGGAAAGTGAACCTCAACACAGTAAACCTGGTACGAGCAGCTTTAGTAAT 718
353 GlnGlnSerAsnPheGlyThrGlyGluLysArgTyrAsnProSerLysThrSerAsn 372
719 GCTAACTTGCATCGTCACTCAAAATGATGAT-----CACAGTGACACA 763
373 GlyHisGlnSerLysSerMetLeuLysAspAspLeuLysLeuSerSerSerGluAspSer 392
764 GATAGTGAAGAGATAGACACAATCAACAGTTTCTCACACTGTAAAGCTTGAATGCA 823
393 AspGlyGluGlnAspCysAspLys-----ThrMetProArgSerThrProGly 408
824 AAGCAGACTCGGAAGATGAACACGACGAGCAAGCCAAAGCCACCAAGTGCAGCAAG 883
409 SerAsnSerGluProSerHisAsnSerGluGlyAlaAspAsnSerArgAspAspSer 428
884 TCTTGCCATCTGGGGAAGACTGTGCAAGTTGTGAGCAAGATGATAGATAGACGTGTGCCA 943
429 SerSerHisSerGlySerGluSerSerSer----- 438
944 AAGAGTCCATTTGTCAGATGTTGGTCTGTGAGAGTGTGTACTGGGTCAAAAATGACAA 1003
439 ---GlySerAspSerGluSerGluSerSerSerSerSerSerSerSerSer 451
1004 AATTTGATTACAGAAAGTTGCTAGCAAAATCTCTCCATTTTGAGAGGAAAGTGA 1063
452 Glu-----AlaAsnGluProSerGlnSerAlaSerProGlu 463
1064 CCCGAATCACCG-----ATGGATGTGGATAAT----- 1090
464 ProGluProProThrAsnLysTrpGlnLeuAspAsnTrpLeuAsnLysValAsnPro 483
1090 ----- 1090
484 HisLysValSerProAlaSerSerValAspSerAsnIleProSerSerGlnGlyTyrLys 503
1091 -----TCTAAAATAGTTCTCAAGACTCAGAAAGCAGATGAGGAG 1129
504 LysGluGlyArgGluGlnGlyThrGlyAsnSerTyrThrAspThrSerGlyProLysGlu 523
1130 ACAACT-----CCAGGTTTTCATGAACAA-----GAAGATGGTAGTTCCTCCCA 1174
524 ThrSerSerAlaThrProGlyArgAspSerLysThrIleGlnLysGlySerGluSerGly 543
1175 ACAGCAATAAACCTTCAAGTTCCACGACGAGCTGACATTTGAATTTAGGAAA--- 1231
544 ArgGlyArgGlnLysSerProAlaGlnSer---AspSerThrThrGlnArgArgThrVal 562
1232 ---CGGTACTCTACTAAGCGCGGTGAAGTTAGATTACATTTTCAATTTGAAGAGGA--- 1285
563 GlyLysLysGlnProLysLysAlaGluLysAlaAlaGluGluProArgGlyGlyLeu 582

```



; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12955  
 ; LENGTH: 1111  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-09-815-242-12955

Alignment Scores:  
 Pred. No.: 9,26e-05 Length: 1111  
 Score: 175.00 Matches: 111  
 Percent Similarity: 34.62% Conservative: 68  
 Best Local Similarity: 21.47% Mismatches: 171  
 Query Match: 2.43% Indels: 167  
 DB: 9 Gaps: 19

US-09-302-812-3 (1-4069) x US-09-815-242-12955 (1-1111)

```

QY 311 AGGTCCTCCACCGTCTCTCGCCAGCC----- 334
DB 203 ArgileProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222
QY 335 ---TGCCTCCAGCGCGGAGCAGCACAGA-----GGCAGCGCCACCTCG 379
DB 223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGlnGlnLysGlnAsnAsp 242
QY 380 CTGTTTTCACAAAGACTATTACCGTGGATGGACACTAAAGGAATCAAGACAGCG 439
DB 243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerThrThr 262
QY 440 GAATCAGAAAGTTTGGATAGTAAGAAACAACTACAGAAATAGAAATCATGATGAGT 499
DB 263 LysSerMetProAsnTyrSerLysValAspAsnThrLysLysLysLysLysLysLys 282
QY 500 TCT-----GTACAAAA----- 511
DB 283 SerGlnLysValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 302
QY 511 ----- 511
DB 303 PheLysLysAlaLeuGlnLysArgGluGluHisLysAsnGluGluGlnAspAlaLe 322
QY 512 -----GATAACTTTTACCACACATAATGTAGAAAAATTTAGTAAATGTTTCTCAG 559
DB 323 GlnArgAlaLeuAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342
QY 560 CTAAGCTTTGATAGTACTACTCTGAAAGAAAGTACACAGTATTTGAACGACATCAGACT 619
DB 343 LeuAsnAspAspSerAspLysThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362
QY 620 GCAGCAATGTGTAAGTGGCAAAATGAAGGAAACACACGACGACGCTTTTGAAGAGTGA 679
DB 363 GlyLe-----Glu 365
QY 680 CTTCAACAGTAAACCTGGTACCAGACGAGTTTAGTAACTGAACATGATCGGTCACT 739
DB 366 AsnGluThrValSer-----AsnAspGluAsnLysGlnAlaSerIle 379
QY 740 CAAATGATGATCAGTACACAGTACATAGTGAAGAGATAGAGACATCAACAGTTTCTC 799
DB 380 GlnAsnGluAspThrAsnAspThrHisLeuAspGluSerProTyrAsnTyrGluGluVal 399
QY 800 ACAACTGTAAAGCTTCAAAATCAAGACGACGACTACGAGAGATGAA-----CACCCAGA 853
DB 400 SerLeuAsnGlnValSerThrThrLysGlnLeuSerAspAspGluValThrValSerAsn 419

```

```

QY 854 GAAGCCAAAGCCACCAAGAGTGCAGCAAGTCTTGCATCTCGGGAAGACTGTGCAAGT 913
DB 420 ValThrSerGlnHisGln-----Ser 426
QY 914 TGTGACAGATGAGATAGACGTGTGTCACCAAGAGTCCATTGTGAGATGTTGGCTCTGAG 973
DB 427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu----- 440
QY 974 GATGTTGGTACTGGGTCAAAATAATGACAAATAATGATTAGACAAAGAAAGTTGCCCTAGA 1033
DB 441 -----LysAsnGlnSerArgLeuLeuAlaAspSerGluLysAspGly 454
QY 1034 AATTCTCTCCATTGAGAGGAAAGTGAACCCGAATCACCGATGATGTGATATATCT 1093
DB 455 AlaThr-----AsnLysGluLysSerGlySerGlnLysAsp-----AspAlaGlu 470
QY 1094 AAAATAGTTTCAAGACTCAGACAGATGAGAGACAGTCCAGGTTTTCATGACAA 1153
DB 471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnLysLysAsn 490
QY 1154 GAAGATGTTAGTCTCTCCCAA-----ACAGCAATAAACCCTTCAAGGTTC----- 1198
DB 491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluTyrAlaVal 510
QY 1199 -----CAAGCAAGACGCTGACATTCGATTAATTTAGGAAA 1231
DB 511 ThrGluSerGlnValAsnAsnLysLysThrValAspAsnGluLysGluLeuAlaPro 530
QY 1232 CGGTAC-----TCTACTAAGCGCGGTCAAGTT 1258
DB 531 ArgHisLysLysAspAspGlnThrAsnLeuSerValAsnSerLysLysThrAsnAspVal 550
QY 1259 AGATTACATTTCCAATTGTAAGAGGAGGAGAGTCCGACTCGAATCAATCAATTAATGCT 1318
DB 551 AsnAspAsnHisValValGluAsp-----SerSerMetAsnGluLysLys 566
QY 1319 AAATACCTGGAATATTTCTAGCTGAATGATGATGATGATGATGATGATGATGATGATG 1378
DB 567 Asn-----AsnAlaGluLysLysLysLysLysLysLysLysLysLysLys 576
QY 1379 AAAAAGGATTTCTAAATCAACATCATTTGATGAGACTGCCCAAGACAGAGACAGAGA 1438
DB 576 AsnGluAlaAlaGluSerGlu-----GlnAsnValGluLys 588
QY 1439 AAAGAACTGCGGAAACCAACATCAAAAGACAGAAAGAGATCCCTAAATACGTTCCA 1498
DB 588 Lys-ThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 608
QY 1499 CTTACCTTTCTCCAGATAGAGTGGCTGGAATCCATCCATCCATCCATCCATCCATCCAT 1558
DB 608 ysArgProPheAsn-----ValValMet-ThrProSerAspLysLysArgMetMet 624
QY 1559 CTTGG-----TGTGGATCGGCTCTCTCTTGGAGACCA 1594
DB 625 AspArgLysLysHisSerLysValAsnValProGluLeuLysPro 639

```

## RESULT 14

US-10-282-122A-43837  
 ; Sequence 43837, Application US/10282122A  
 ; Publication No, US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.

1 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

2 FILE REFERENCE: ELITRA.034A

3 CURRENT APPLICATION NUMBER: US/10/282,122A

4 CURRENT FILING DATE: 2003-02-20

5 PRIOR APPLICATION NUMBER: 60/191,078

6 PRIOR FILING DATE: 2000-03-21

7 PRIOR APPLICATION NUMBER: 60/206,848

8 PRIOR FILING DATE: 2000-05-23

9 PRIOR APPLICATION NUMBER: 60/207,727

10 PRIOR FILING DATE: 2000-05-26

11 PRIOR APPLICATION NUMBER: 60/230,335

12 PRIOR FILING DATE: 2000-09-06

13 PRIOR APPLICATION NUMBER: 60/230,347

14 PRIOR FILING DATE: 2000-09-09

15 PRIOR APPLICATION NUMBER: 60/242,578

16 PRIOR FILING DATE: 2000-10-23

17 PRIOR APPLICATION NUMBER: 60/253,625

18 PRIOR FILING DATE: 2000-11-27

19 PRIOR APPLICATION NUMBER: 60/257,931

20 PRIOR FILING DATE: 2000-12-22

21 PRIOR APPLICATION NUMBER: 60/267,636

22 PRIOR FILING DATE: 2001-02-09

23 PRIOR APPLICATION NUMBER: 60/269,308

24 PRIOR FILING DATE: 2001-02-16

25 Remaining Prior Application data removed - See File Wrapper or PALM.

26 NUMBER OF SEQ ID NOS: 78614

27 SOFTWARE: PatentIn version 3.1

28 SEQ ID NO 43837

29 LENGTH: 1274

30 TYPE: PRT

31 ORGANISM: Staphylococcus aureus

32 US-10-282-122A-43837

Alignment Scores:

Pred. No.: 0.000118 Length: 1274  
Score: 174.00 Matches: 112  
Percent Similarity: 35.28% Conservative: 69  
Best Local Similarity: 21.83% Mismatches: 173  
Query Match: 2.41% Indels: 159  
DB: 12 Gaps: 20

US-09-302-812-3 (1-4069) x US-10-282-122A-43837 (1-1274)

QY 311 AGGCTCCACGCTCTCCGACGCC----- 334  
DB 203 ArgileProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222  
QY 335 ---TGGCTCCAGGCGGCGGACACACAGA-----GGCAGCGCCACCTCG 379  
DB 223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGluGlnGluLysGlnAsnAsp 242  
QY 380 CTGTGTTTCAACAAACAGACTATTACCAAGTTGGATGGACATAAAGGAATCAAGACACGCG 439  
DB 243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerSerThrThr 262  
QY 440 GAATCAGAAAGTTGGATAGTAAGAAACAAACATAACAGATAGATCCATGATGAGT 499  
DB 263 LysSerMetProAsnTyrSerLysValAspAsnThrLysLysLysLysLysLysLysLysLys 282  
QY 500 TCT-----GTACAAAA----- 511  
DB 283 SerGlnIleValGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 302  
QY 511 ----- 511  
DB 303 PheLysLysAlaLeuGlnGlnLysArgGluGluHisLysAsnGluGluGluAspAlaIle 322  
QY 512 -----GATAACTTTTACCAACATAATAGTAAATAATAGTAAATTTCTCAG 559  
DB 323 GlnArgAlaIleAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342  
QY 560 CTAAGTCTGTAAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 619

Db 343 LeuAsnAspAspSerAspLeuThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362  
QY 620 GCAGCAATGTGTAAGTGCAGAAATCAAGGAAACACACGGACGACTTTTGGAAAGTAA 679  
Db 363 GlyIle-----Glu 365  
QY 680 CCTCAACACAGTAACCTGGTACCAGAGCAGTTTACTAATCTTAACATTGATCGGTCACT 739  
Db 366 AsnGluThrValSer-----AsnAspGluAsnLysGlnAlaSerIle 379  
QY 740 CAAATATGATGATCAGACAGTACAGATAGTGAAGAGATAGAGACAAATCAACAGTTCTC 799  
Db 380 GlnAsnGluAspThrAsnAspThrHisValAspGluSerProTyrAsnTyrGlnGluVal 399  
QY 800 ACAACTGTAAAGCTGCAAAATCAAGACAGACTACGGAAGATCAA-----CACGCCAGA 853  
Db 400 SerLeuAsnGlnValSerThrThrLysGlnLeuSerAspAspGluValThrValSerAsn 419  
QY 854 GAAGCCAAAAGCCACAGAAAGTGCAGCAAGTCTTGCATCTCTGGGGAAGACTGTGCAAGT 913  
Db 420 ValThrSerGlnHisGln-----Ser 426  
QY 914 TGTCAACAAGATGAGATAGACAGCTGGTCCAAAGAGTCCATTGTTCAGATGTTGCTCTGAG 973  
Db 427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu----- 440  
QY 974 GATGTTGGTACTGGGTCAAAAATGACAAATTCATTAGACAAAGAAAGTTGCCTTAGGA 1033  
Db 441 -----LysAsnGlnSerArgLeuIleAlaAspSerGluGluAspGly 454  
QY 1034 AATTCTCTCCATTTCAGAAAGAAAGTGAACCCGAATCACCGATGGATGGATAATTCT 1093  
Db 455 AlaThr-----AsnLysGluGluTyrSerGlySerGlnIleAsp---AspAlaGlu 470  
QY 1094 AAAAATAGTTGTCAAGACTCAGAAAGCAGATGAGAGACAGTCCAGTCTGATGATGACAA 1153  
Db 471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnIleGluAspAsn 490  
QY 1154 GAAGATGGTGTCTCTCCCAACAGCAAAATACCTTCAAGGTTCCAAAGCAAGA----- 1207  
Db 491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluHisAlaVal 510  
QY 1208 -----GACCTGCATGATTAATTTAGGAAA 1231  
Db 511 ThrGluSerGlnValAsnAsnIleAspLysThrValAspAsnGluIleGluLeuAlaPro 530  
QY 1232 CGGTACTCTACTAAGCGCGGTGAAGTTAGATTACATTTCCATTTGAAGGAGGAGAGT 1291  
Db 531 ArgHisLys---LysAspAspGlnThrAsnLeuAsnVal-----Asn 543  
QY 1292 CGCACTGGAATGATGATTTAAATGCTAAACTACCTGCAATATT-----TCTAGC 1342  
Db 544 SerLeuLysThrAsnAspValAsn-----AspGlyHisValValGluAspSerSer 560  
QY 1343 CTGAAT-----GTAGATGCGAGAAATTTCTAAGCAATCATGAGAAAAGAGTTCT 1390  
Db 561 MetAsnGluIleGluLysHis-AsnAlaGluIleThrGluAsnValGlnAsnGluAlaIle 580  
QY 1391 AAAATCAGACATCATTTGATGAGACTCCCAAGCAGGACAGCAAGAAAGACAGTGG 1450  
Db 580 aGluSerGlu-----GlnAsnValGluGluLys-ThrIleG 592  
QY 1451 GAAACCAACATCAACAGCAAGAAAGAGTCCCTAAATACGTTCCACCTTCACCTTTCT 1510  
Db 592 luAsnValAsnProLysGlnThrGluLysValSerThrLeuSerLysArgProPheA 612  
QY 1511 CCAGATAAGAGTGGCTTGGAACTCCCATTTGAGAGATGAGAAAGATGCCTCGG----- 1564  
Db 612 sn-----ValValMet-ThrProSerAspLysLysArgMetMetAspArgLysLys 628  
QY 1565 ---TGTGGATCCGCGTCTCTCTTGTAGACCA 1594  
Db 629 HisSerLysValAsnValProGluLeuLysPro 639





1250	QY	GGTGAAGTTAGATTACATTTCCAATTT	1276
346	Db	SerGlnLysLysThrSerLeuTyrPheAspGluGluProGluLeuGluMetGluSerLeu	365
1277	QY	GAAGGAGGAGAGAGTGCACAT	1321
366	Db	ThrAspSerProGluAspArgSerArgGlyGluGlySerSerLeuHisAlaSerSer	385
1322	QY	CTACCTCGAAATATTTCTTAGCCTGAATTCAGATTCAGAAATCTT	1366
386	Db	PheThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSerAspSer	405
1367	QY	AAGCAACATCGAAAAAGGATTCATAA	1408
406	Db	ProSerHisLysLysGlyGluSerLysGlnArgLysAlaArgHisArgProHisGly	425
1409	QY	ATGAGATTCGCCAAAGCAGAGACAGA	1450
426	Db	ProLeuLeuProThrIleGluAspSerSerGluGluGluLeuArgGluGluGlu	445
1451	QY	GAACCAACATCAAAGAACAGAAAGCAATC	1483
446	Db	LeuLeuLysGluGlnGluLysGlnArgGluLeuGlnGlnArgLysSerSer	465
1484	QY	CTATAA	1489
466	Db	LysLysSerLysLysAspLysAspGluLeuArgAlaGlnArgArgGluArgProLys	485
1490	QY	TAGTTCCACTCACCTTCTCCAGATAAGAGTGGCTTGGAACTCCATTCAGGAGATG	1549
486	Db	ThrProProSerAsnLeuSerProIleGluAsp	503
1550	QY	AGAAAGATGCTCGGTGGGATCCGGCTGCTCTTTGAGACCATCT	1606
504	Db	ArgGlnAlaAlaGluMetGlu	520
1607	QY	ACAGTAACATTCGGGTAGATCTTTTCGGAGGAGGAGAGTCTCTAAACCTTTCCAC	1666
521	Db	SerProSerIleGluSerAsp	533
1667	QY	CATTATAAGATTGTGGAT	1723
534	Db	SerProGluLysIleIleGluValGlnLysValTyrLysLeuPro	552
1724	QY	TTGTACCACGAGTGAAGATGAGATGGTGAGCAACTCGGGGAGCCGGTGGGAGCTCAT	1783
553	Db	LeuTyrSerProThrAspGluGln	560
1784	QY	CAGACTGCATCTCAACAATTTACAGACCCCAAACTTGAAGATGCTATCTGAAA	1843
561	Db	SerIleMetGlnLysGluGlySerGlnLysAlaLeuLysSerAlaGluGluMet	578
1844	QY	TACAATGGCATATCTTAAGAAATGGACTTTACAGCTTG	1897
579	Db	TyrGluGluMetMetHisLysThrHisLysTyrLysAlaPheProAlaAlaAsnGluArg	598
1898	QY	GATAGGTACTTTGAAGACAGCAA	1936
599	Db	AspGluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAspTyrIleTyr	618
1937	QY	CAGTCCATCTTGGCTGAT	1954
619	Db	GluSerLeuValGluAsp	624



QY 1988 -----TGCACCCAGCCCAATACCACTCTCTGAAACAGAGATGAATCATTC 2032  
 Db 119 AsnIleValSerGlyIleLeuThrGlyLeuArgLeuLeuAenSerGlnAlaGlyIle 138  
 QY 2033 ATCAATGTCGAGAACAGATGCCAGTCTTTTAGCTAATGCTTTCTTCTGACATTT 2092  
 Db 139 ValPheLeuSerGlnGluLeuIleGlyAlaLeuLeuAlaCysSerPhePheCysLeuPhe 158  
 QY 2093 CCACGACGAATGCTAGATCAAAATCGGAGTATTCTAGTATCCACGACATTAACATCAAT 2152  
 Db 159 ProAspAsnArgGlyAlaLys-----HisLeuProValIleAsnPheAsp 174  
 QY 2153 CGA-----TTGTTTGGGGACGTTTCATCAAGAAACCGGAGAACTTAAACGCTTTC 2206  
 Db 175 HisLeuSerLeuTyrIleSerTyrSerGlnSerGlnGluSerLysIleArgCysIleMet 194  
 QY 2207 TGCTACTTTAGAGAGTACAGAGAAAACCTACTGGTGGTGCACATTTACAGACAG 2266  
 Db 195 HisTyrPheGluArgPheCysSerCysValProIleGlyIleValSerPheGluArgLys 214  
 QY 2267 -----AGTCTTGAAGATTTTCCAGAAATGGGAAGATGTGAAACCCCTTCACACGATTG 2320  
 Db 215 IleThrAlaAlaProAspAlaAspPheTyrSerLysSerAspValSerLeu----- 231  
 QY 2321 CATCTCACTTACGAAAGTACCATAGAAAGAAATGGCCAAAGCATGCTACAGGTGATTTT 2380  
 Db 232 -----TyrGln-----ProAspAsnAlaLeuGluValAspPhe 242  
 QY 2381 GCAATCGTTTGTGGAGGTGGTCTAACCATGTCAGGACTTGTGCAAGAGAAATCCGC 2440  
 Db 243 AlaAsnLysTyrLeuGlyGlySerLeuSerArgGlyCysValGlnGluGluLeuArg 262  
 QY 2441 TTTTAAATCAATCTGAGTCTGATTTTCAAGGTCTTTCAGGTCTCTGCTGCTGACACAA 2500  
 Db 263 PheMetIleAsnProGluLeuIleAlaGlyMetLeuPheLeuProArgMetAspAsn 282  
 QY 2501 GAATGCTAAATATACAGGTACTGAGCAGTACAGTGAATACACAGGCTTACTGAGACA 2560  
 Db 283 GluAlaIleGluIleValGlyAlaGluArgPheSerCysTyrThrGlyTyrAlaSerSer 302  
 QY 2561 TATCGTTGGTCCGAGCCACGAGATGGAGTGAAGGGACGACTGCGAGCGGGCTGTC 2620  
 Db 303 PheArgPheAlaGlyGlyTyrIleAspLysAlaMetAspProPheLysArgArg 322  
 QY 2621 ACTGAGATCGTTGCCATGCTCTT-----CACTTCAGACGCTAC 2662  
 Db 323 ThrArgIleValAlaIleAspAlaLeuCysThrProLysMetArgHisPheLysAspIle 342  
 QY 2663 CTCGATCAGTTTGTCCCTGAGAAATGACACGAGCTGACACAGGCTTACTGATTT 2722  
 Db 343 Cys-----LeuLeuArgGluIleAsnLysAlaLeuCysGlyPhe 355  
 QY 2723 CTCGGT----- 2728  
 Db 356 LeuAsnCysSerLysAlaTyrGluHisGlnAsnIlePheMetAspGluGlyAspAsnGlu 375  
 QY 2728 ----- 2728  
 Db 376 IleGlnLeuValArgAsnGlyArgAspSerGlyLeuLeuArgThrGluThrAlaSer 395  
 QY 2728 ----- 2728  
 Db 396 HisArgThrProLeuAsnAspValGluMetAsnArgGluLysProAlaAsnLeuIle 415  
 QY 2729 -----CCTGGAGTTTCTTCAGAGAATCTT-----TCTGACGTGCCACA 2767  
 Db 416 ArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspAspGlyValAlaThr 435  
 QY 2768 GGAACCTGGGGCTGTGGTCCCTTGGGGGTGATGCCAGGTAAAGCCCTTAATACAGATA 2827  
 Db 436 GlyAsnTyrGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIleGlnTyr 455

QY 2828 TTGGCAGCTCTGCGAGCTGAGCGAGATGTGGTT---TATTTCACCTTTGGGACTCAGAA 2884  
 Db 456 LeuAlaIleSerGlnThrArgArgProPheIleSerTyrThrPheGlyValGlu--- 474  
 QY 2885 TTGATGAGACACATTTACAGCATGCACATTTCTCTACTGAAAGAAACACTCACTGTGGA 2944  
 Db 474 ----- 474  
 QY 2945 GATGTGTATAAGCTGTGTGCTACGATCTCAATGAGAAGTGCAGAACTGTTCCACCC-- 3002  
 Db 475 -----AlaLeuArgAsnLeuAspGlnLeuProProTh 485  
 QY 3003 -----CTGACACGACATCAAGC-----TTTATCCATTC 3031  
 Db 485 rLysLeuSerSerArgLeuAspSerAlaAsnSerProSerLeuThrCysLeuMetHisSe 505  
 QY 3032 ATATACCATGCTGTGAGTCTGTGCGAGACCGCTCACCATTCCAGGGCAAGACAGGG 3091  
 Db 505 rHisThr-----ValSerLeuHisPhe-ProThrTyrAlaIleArgMetG 520  
 QY 3092 ACTGAGGAGCGGAGCGAATAGCATCTCTCCACCTCCCA-----CCAGAGACGCTCT 3145  
 Db 520 LuLeuArgAlaAspLeuArgSerIleLeuGlnTyrLeuProLeuValAlaGlnSerSerS 540  
 QY 3146 GTTTGAGCTGTGAGTGTATATATGAATTGACTTAAAGTTAATATAATGTGTACATAAT 3205  
 Db 540 erLeuValTyrProSerValGluGluGluLeuGlnThrIleSerArgGlyProSerG 560  
 QY 3206 CCACATTTGTAGTCAAGAGCGCAATCTCTTCCACACATGTG-----CAGT 3250  
 Db 560 LuSerMetValAsnSerGlyGluAlaLeuAlaLeuHisIleThrAsnMetArgLysSerL 580  
 QY 3251 TGTGAGTTGTACATCTAAACTCCCTCCATCTCAGCTCAGCTGACCTAGATATGTTTG 3310  
 Db 580 euSerLeuAsnAlaSerAspLeuAlaProTyrAlaLeuGingly-----TyrGlyLeu 597  
 QY 3311 TTTCTATTTTCTTCTATTCTA-----GTTTTTCAT 3340  
 Db 598 PhePheAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValPro 617  
 QY 3341 TCTTTGATGTTTATTTCTTTTGTCCATCAGATCTCTGTGTAATCCCAT 3389  
 Db 618 AlaLeu-CysArgLeuLeuGlnLeuProSerMetLeuGluLysHis 633  
 RESULT 2  
 T21138  
 hypothetical protein F20C5.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T21138  
 R:Matthews, P.  
 submitted to the EMBL Data Library, December 1995  
 A:Reference number: Z19381  
 A:Accession: T21138  
 A:Status: preliminary; translated from GB/EMBL/DBBJ  
 A:Molecule type: DNA  
 A:Residues: 1-726 <WIL>  
 A:Cross-references: EMBL:Z68161; PDB:CAA922299.1; GSPDB:GN00022; CESP:F20C5.1  
 A:Experimental source: clone F20C5  
 C:Genetics:  
 A:Gene: CESP:F20C5.1  
 A:Map position: 4  
 A:Intron: 15/3; 162/3; 219/3; 294/1; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1  
 Alignment Scores:  
 Pred. No.: 4.5e-25 Length: 726  
 Score: 463.00 Matches: 207  
 Percent Similarity: 38.11% Conservative: 123  
 Best Local Similarity: 23.90% Mismatches: 278  
 Query Match: 6.42% Indels: 258  
 Gaps: 37  
 DB: 2



T34036  
hypothetical protein B0041.7 - *Caenorhabditis elegans*

	::	:::	:::			:::		ProSerLysLysSerIysl	198
Df	178	yEargAlaGluThrSerGluGlusErAspGluaspGluLysProSerLysLysSerIysl							
QY	792	AGTTTCTCAACACTGTAAAGCTTGCAAATTCGAAGCAGACTACGGAAGATGAACAAGCCCA	851						
Df	198	yGlyLeuLys-----LysLysAlaLysSerGluSerGluSerGluaspGluL	216						
QY	852	GAGAACCAAAACGCCACAGAGTCAGCAGAGTCTTGCCATCTCTGGGAGACTGTGC	911						
Df	216	ySgluValLysLysSerLysLysLysValLysLysGluSerGluSerGluSerG	236						
QY	912	GTGTGCAGCAAGATGATAGACGTGTCGCAAGAGTCCATTGTAGATGTTGGTCTCG	971						
Df	236	luaspGluAlaProGluLysLysThrGluLysArgSerLysThrSerSerG	256						
QY	972	AGGATGTTGTAAGTCTCAAAAAATGCAACAAATGATGATAGCAGAGAAGTTCCTAG	1031						
Df	256	luGluSerSerGluSerGluLysSerAspGluLys-----GluGlu	269						
QY	1032	GAATTTCTCTCATTTGAGAAGGAAGTCAACCC							
Df	270	-----GluLysGluSerSerProLysProLysLysLysLysProLys	284						
QY	1080	ATGTGATATAATCTAAAAATAGTTGTCAAGACTCAGAACGAGATGAGAGCAAGTCCA-	1138						
Df	284	laValLysLysLeuSerSerAspGluGluSerGluSerAspValGluValLeuProG	304						
QY	1139	-----GGTTTTGATGAACAAGAAGATG	1160						
Df	304	lnLysLysLysArgLysAlaValThrLeuLysSerAspSerGluaspGluLysaspGluL	324						
QY	1161	GTAGTTCTCCCAACAGCAACAAATAAACCCTTCAAGGTTCCAAAGCAGAGACGCTCACAT	1220						
Df	324	ySerGluSerGluAlaSerAspValGluLysValSerLysLysLysala	341						
QY	1221	AATTTAGAAACGGTACTCTACTAAGGGCGGTGAAGTAGATTACATTTCCAATTTGAAG	1280						
Df	342	-----LysLysGluGluSerSerGlu	350						
QY	1281	GAGCAGAGACTCGACTGGGAATGAATGATTAAATGCTAACTACCTCGGAATATTTCTA	1340						
Df	350	lysErAspSerSerGluGluLysErLleThrValasnArgLys	364						
QY	1341	GCCTGAATGTAGATGCAGAAATTTCTAAGCAACATGGA	1394						
Df	364	erLysLysLysGluLysProGluLysLysLysLysGlyLlleMetAspSerSerLysL	384						
QY	1395	TCAGATCATTTGATGAGACTGCCAAAGCAGAGGACAGAGAAGAAAGACAGTGGGAAA	1454						
Df	384	eugLnLysGluThrIleAspAlaGluargAlaGluLysGluargLysargLeuGluL	404						
QY	1455	CCAAACATCAAAGA-----	1468						
Df	404	ysLysGlnLysGluPheasnGlylileValLeuGluGluGlyLysLeuThrGluMetL	424						
QY	1469	-----ACAGAAAGAGATCCCTAAATACGTTCCACCTCACCTTTCTCCAG	1514						
Df	424	euthrGlyThrSerSerGlnArgLysLeuLysSerValVal-----LeuaspProA	441						
QY	1515	ATAAGAAGTGGCTTGGAACTCCCATTGAGAGATGAGAGAATGCCTGGTGTGGGATCC	1574						
Df	441	sp-----SerSerThrValaspGluGluSerLysLysPro-----	452						
QY	1575	GGCTGCCTCTCTGAGACCATCTGCCAATCACACAGTAACTATTTCGGGTAGATCTTTG	1634						
Df	453	-----ValGluValHisasnSerLeuValA	461						
QY	1635	GAGCAGGAGAGTTCCTAAACCTTTTCCACACATTTATAAGATTTGTGGGATAACAAGC	1694						
Df	461	rg-----IleleuLysProHisGlnAlaHIs	469						
QY	1695	ATGTTAAAAATGCCTTGTTCAGAACCAAAATTTGTATCCCATGTCGGAAGATGAGATG	1754						

```

Db 470 -----GlylleGlnPheMetTyrAspCysAlaPheGluSerLeuAspA 484
QY 1755 GA-----ACTGGGGAGCGGTGGAGCTCATTGAGCTGCTCTCAACAATTTA 1808
Db 484 rgleuaspThrGluGlySer-----490
QY 1809 CAGACCCCAAACTGAAGAGTCTATTCTGAAATACATATGTCGATATTAAGAAAT 1868
Db 491 -----GlyGlylleLeuAlaHisCysMetGlyLeuGlyIysThr- 503
QY 1869 GGGACTTTACAGCTTTTCGATTCGATTCGGGATGAGTACTTGAAGACGAGAAGCTCAAC 1928
Db 504 -----LeuGlnValIleThrPheLeuHisThrValLeu-----514
QY 1929 ATTATATCATCATCTTCCTGATATGTTGGTAAATATGCACTCTGTCGCCA-----ATA 1985
Db 515 -----MetHisGluLysIleGlyGluLysCysLysArgValLeuValValProLysAsnV 534
QY 1986 TTTGCCACCCAGCCATACCCTCTGAAACAGAGATGAATCATTCATCACAATGTGCG 2045
Db 534 alileileAsnTrpPhe-----LysGluPheGlnLysTrpLeuValAspAsnA 550
QY 2046 AGGAACAGATTCGCCAGTCTTTAGCTAATGCTTTCTTCGCACATTTCCACGAGAAATG 2105
Db 550 spGluGluLeuAspThrIleAspValAsn-----559
QY 2106 CTAAGATGAATCGAGTATTCATGATCCAGACATTAATCTCAATCGA-----TTGT 2159
Db 560 -----GluLeuAspSerTyrLysThrIleGluAspArgArgAlaLeu 575
QY 2160 TTGAGGACGCTTCATCAGGAAACCGGAGAACTTAAACGCTCTCTGCTACTTTAGAA 2219
Db 575 ysAlaTrpHisSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgI 595
QY 2220 GAGTCACA-----GAGAAAAACCTACTGCTGGTGGTGGTACATTTACAA 2261
Db 595 leLeuThrValGluAspAspProLysLysLysLysProLysAsnArgAsnArgLeuG 615
QY 2262 GACGAGCTCTTGAGATTTT-----CCGAAATGGGAAGAT 2297
Db 615 luLysAlaLysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValC 635
QY 2298 GTGAA-----AAAC 2306
Db 635 ysAspGluAlaHisLysLysLysAsnAspAspSerAlaLeuSerLysCysMetValLysI 655
QY 2307 CTTGACACGATTCATGCTCACTTACGAGGTACCATAGAGAAATGGC-----2356
Db 655 leLeuThrLysArgArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluT 675
QY 2357 -----CAAGGCATGCTA-----CAGGTGGATTTTGCAA 2384
Db 675 yrHisCysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaA 695
QY 2385 ATCGTTTGTGGAGGTGGTAAACGACGTCGAGGACTTGTGCAAGAA-----G 2432
Db 695 snArgPheVal-----AsnIleileAsnArgGlyArgThrLysAspAlaSerProLeuG 713
QY 2433 AAATCCGCTTTTAAATCAATCTGAGTGGATTTATTCACGGCTCTTCACTAGGTGCTGG 2492
Db 713 luValSerPheMetLysArgArgCysHisValLeuTyrAspHisLysLysCysValA 733
QY 2493 ATCACAATGAATGCTAATTATCACAAGTACT-----2524
Db 733 spArgLysAspTyrArgValLeuThrGluAlaIleProProLysGlnGluTyrValIleA 753
QY 2525 -----GAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGTT 2567
Db 753 snValArgGlnThrGluArgGlnCysAlaLeuTyr-AsnAlaPheLeuAsnAspIleVal 772
QY 2568 GGTCCCGGAGCCACGAAGATGGAGTGAAGGACGACGTCGCGAGCGGCTCGACTGAGA 2627
Db 773 GlyAsp-----SerGlyLeuSerLysArg 780

```

```

QY 2628 TCGTTGCCA 2636
Db 781 LeuLeuPro 783

RESULT 5
154367
X-linked nuclear protein - human
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I54367
R;Gecz, J.; Pollard, H.; Consalez, G.; Villard, L.; Stayton, C.; Millasseau, P.; Khrestc
Hum. Mol. Genet. 3, 39-44, 1994
A;Title: Cloning and expression of the murine homologue of a putative human X-linked nuc
A;Reference number: I54367; MUID:94214473; PMID:8162050
A;Accession: I54367
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1298 <RES>
A;Cross-references: GB:I34363; NID:g508653; PIDN:AAA20872.1; PID:g530788
C;Genetics:
A;Gene: XNP

Alignment Scores:
Pred. No.: 0.000107 Length: 1298
Score: 180.00 Matches: 199
Percent Similarity: 34.89% Conservative: 160
Best Local Similarity: 19.34% Mismatches: 374
Query Match: 2.50% Indels: 296
DB: 2 Gaps: 53

US-09-302-812-3 (1-4069) x I54367 (1-1298)
QY 389 AAACAAAGACTATTACAGTGGATGGACACATAAGAAATCAAGACAGCGGAATCAGAA 448
Db 17 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluIleGlnSerGlySerSer 36
QY 449 AGTTTGGATAGTAAAGAAAC-----ACAATCAAGAATAGAAATCAATGATGATCTCT 502
Db 37 Ser---AspAlaGluGluSerSerGluAsnLysLysLysGlnArgThrSerSer 55
QY 503 GTACAAAAGACTATTTCACACATAATAGTAAAAATTTAGTAAATTTCTTCAGCTA 562
Db 56 LysLysLysAlaValIleValLysGluLysLysArgAsnSerLeuArgThrSerThrLys 75
QY 563 AGTCTTGATAGTCACTCACTCACTAAAAAGTACACAGTATTTTGAACACGACATCAGCTGCA 622
Db 76 ArgLysGlnAlaAspIleThrSerSerSerSerSerSerSerSerSerSerSerSerSer 95
QY 623 GCAATGTGTAAAGTGGCAAAATGAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCT 682
Db 96 SerIle-----GlyGluGlySerSerSerSerSerSerSerSerSerSerSerSer 109
QY 683 CAACAGTAACCCCTGGTACCAGAGCAG-----TTTACTAATGCTTAACATTCATCGG 733
Db 110 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 129
QY 734 TCA-----CCTCAAAATGATGATCATCAGTCAGTCACACAGATAGTGAA 772
Db 130 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAspAspAsp 149
QY 773 GAGAAATAGAGCAATCAACAGTCTTCTCAACTGTAAAG-----CTTGCAAAATGCAAG 826
Db 150 AsnArgIleAlaLysLysMetLeuLeuGluLysLysLysLysLysLysLysLysLys 169
QY 827 CAGACTACGGAAGTGAACACGCGCAGAGAGCCAAAGCCACCAAGTGCAGCAGTCT 886
Db 170 AspGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 189
QY 887 TGCCATCTCTGGGAGACTGTGCAAGTTGTGACGAAAGATGATAGATAGCTGGTGCCAAAG 946
Db 190 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer-----203

```





Db 813 ValGluValIleLysValTrpAsnSerArgSerArgGlyGlyGlyGlu----- 828  
 Qy 2816 TTAATACAGATATTGGCAGCTGCTGCAGCTGAGCAGATGGTTATTTCACCTTTGGG 2875  
 Db 829 -----Gly 829  
 Qy 2876 GACTCAGATTGTAGAGACATTTACAGCATGCACATTTTCCTTACTGAAAGGAACTC 2935  
 Db 830 AsnValAspGluThrGlyAsnAsnProSerValSerLeuLysLeuGluGluSerLysAla 849  
 Qy 2936 ACTGTTGGAGATGTTATAGCTGTTCTACCATPACTACATGAGAAATGCGAAGTGT 2995  
 Db 850 Thr-----SerSerSerAsnPro 855  
 Qy 2996 TCACCCCTGGACACAGACATCAAGCTTTATCCATTATATACCATGCTGCTGCTGCTGT 3055  
 Db 856 SerSerProAlaProAspTrp-----TyrLysAspPheValThrAspAlaAsp 871  
 Qy 3056 GCAGAGACCGCTGACCATTCAGGGCAA 3082  
 Db 872 AlaGluValLeuGluHisSerGlyLys 880  
 RESULT 6  
 A:89959  
 hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 R:Accession: AB9959  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: AB9758; MUID:21311952; PMID:11418146  
 A:Accession: AB9959  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1274 <KUR>  
 A:Cross-references: GB:BA000018; PID:gl3701536; PIDN:BAB42830.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA1562

## Alignment Scores:

Pred. No.:	0.000287	Length:	1274
Score:	174.00	Matches:	112
Percent Similarity:	35.28%	Conservative:	69
Best Local Similarity:	21.83%	Mismatches:	173
Query Match:	2.41%	Indels:	159
DB:	2	Gaps:	20

US-09-302-812-3 (1-4069) x AB9959 (1-1274)  
 Qy 311 AGGTCCTCACCGTCTCTCCGACGC----- 334  
 Db 203 ArgileProValSerLysProSerGluLysValGluSerAspLysGlnLysTyrAspLys 222  
 Qy 335 ---TGCCTCCAGGCGCGGAGCAGCAGCAG-----GGCAGCGCCACCTCG 379  
 Db 223 TyrValAlaLysThrGlnThrSerGlnAsnLysGlnLeuGluGlnGluLysGlnAsnAsp 242  
 Qy 380 CTGTGTTTCAACAAAGACATATTACCATGTTGGATGGACACTAAAGGAATCAAGACAGC 439  
 Db 243 SerValValLysGlnGlyThrAlaSerLysSerSerAspGluAsnValSerSerThrThr 262  
 Qy 440 GAATCAGAAAGTTGGATAGTAAAGAAACAACAATACAGATAGATCCATGATGAGT 499  
 Db 263 LysSerMetProAsnTyrSerLysValAspAsnThrLysLysLysLysLysLysLysLys 282  
 Qy 500 TCT-----GTACAAAA----- 511  
 Db 283 SerGlnIleValGluLysValArgGluArgGluArgLysValIleGlnLysArgArg 302

Qy 511 ----- 511  
 Db 303 PheLysLysAlaLeuGlnGlnLysArgGluGluHisLysAsnGluGlnAspAlaIle 322  
 Qy 512 -----CATAACTTTTACCAACATATGTAGAAAAATAGTAAATTTTCTCAG 559  
 Db 323 GluArgAlaIleAspGluMetTyrAlaLysGlnAlaGluArgTyrValGlyAspSerSer 342  
 Qy 560 CTAAGTCTTCATAGTCACTCACTGAAAGAAATACACAGTATTGTGACCAACGATCAGACT 619  
 Db 343 LeuAsnAspSerAspLeuThrAspAsnSerThrAspAlaSerGlnLeuHisThrAsn 362  
 Qy 620 GCAGCAATGTGTAGTGGCAAAATGAAGGAAACACACGAGCAGCTTTTGGAAAGTGAA 679  
 Db 363 GlyIle-----Glu 365  
 Qy 680 CCTCAACAGTAACCTGGTACCAGACGATTTAGTAATGCTAACATTTGATCGTCACT 739  
 Db 366 AsnGluThrValSer-----AsnAspGluAsnLysGlnAlaSerIle 379  
 Qy 740 CAAATGATGATCACAGTCACACAGATAGTGAAGAGAAATAGAGACAATCAACAGTTTCT 799  
 Db 380 GluAsnGluAspThrAsnAspThrHisValAspSerProTyrAsnTyrGluGluVal 399  
 Qy 800 ACAACTGTAAGCTTGCAPATGCAAGACGACACTACGGAAGATGAA-----CACGCCAGA 853  
 Db 400 SerLeuAsnGlnValSerThrLysGlnLeuSerAspAspGluValThrValSerAsn 419  
 Qy 854 GAAGCCAAAGCCACCAACAGAGTGCAGCAAGTCTTGCCATCTCGGGAAGACTGTGCAAGT 913  
 Db 420 ValThrSerGlnHisGln-----Ser 426  
 Qy 914 TGTCAAGACAGATGAGATAGACAGTGGTGCACAAAGAGTCCATTGTGCAGATGTTGGCTCTGAG 973  
 Db 427 AlaLeuGlnHisAsnValGluValAsnAspLysAspGluLeu----- 440  
 Qy 974 GATGTTGGTACTGGGTCAAAAATAGCAACAATGATTAGTACACAGAAAGTTGCCTAGGA 1033  
 Db 441 -----LysAsnGlnSerArgLeuIleAlaAspSerGluGluAspGly 454  
 Qy 1034 AATTCCTCTCCATTGGAAGGAAGTGAACCCGAATCACCAGTGGATGTGGATAATTCT 1093  
 Db 455 AlaThr-----AsnLysGluGluTyrSerGlySerGlnIleAsp---AspAlaGlu 470  
 Qy 1094 AAAAATAGTTGTCAAGACTCAGACAGCAGATGAGAGACAAGTCCAGGTTTGTATGAACA 1153  
 Db 471 PheTyrGluLeuAsnAspThrGluValAspGluAspThrThrSerAsnIleGluAspAsn 490  
 Qy 1154 GAAGATGGTGTCTCTCCAAACAGCAATAAACCTTCAGGTTCCAGCAAGA----- 1207  
 Db 491 ThrAsnArgAsnAlaSerGluMetHisValAspAlaProLysThrGlnGluHisAlaVal 510  
 Qy 1208 -----GACGCTGACATTTGAATTTAGGAAA 1231  
 Db 511 ThrGluSerGlnValAsnAsnIleAspLysThrValAspAsnGluIleGluLeuAlaPro 530  
 Qy 1232 CGGTACTCTACTAAGCGGTGAAGTTAGATTACATTCCATTTCATTTGAGGAGGAGAGT 1291  
 Db 531 ArgHisLys-----LysAspAspGlnThrAsnLeuAsnVal-----Asn 543  
 Qy 1292 CGCACTGGATGAATGATTAAATGCTAAACTACCTGCAATATT-----TCTAGC 1342  
 Db 544 SerLeuLysThrAsnAspValAsn-----AspGlyHisValValGluAspSerSer 560  
 Qy 1343 CTGAAT-----GTAGAATGCAGAAATTTAAGCAACATCGGAAAAAGAGTTCT 1390  
 Db 561 MetAsnGluIleGluLysHis-AsnAlaGluIleThrGluAsnValGlnAsnGluAlaAl 580  
 Qy 1391 AAAATCAGACATCTTTGATGAGACTGCCAAAGCAGGAGCAGACAGAAAGACAGTGG 1450  
 Db 580 aGluSerGlu-----GlnAsnValGluGluLys-ThrIleG 592  
 Qy 1451 GAACCAACATCAAGACAGAAAGAGATCCCTAAATACGTTCCACCTCACCTTCT 1510

```

Db      592 luAnValAsnProLysGlnThrGluLysValSerThrLeuSerLysArgProPheA 612
QY      1511 CCAGATAGAGTGGCTTGGAACTCCCATTCAGGAGATGAGAGAAATCCCTCGG----- 1564
Db      612 sn-----ValValMet-ThrProSerAspLysArgMetMetAspArgLysLys 628
QY      1565 ---TGTGGATCGGCTGCTCTCTTTGAGACCA 1594
Db      629 HisSerLysValAsnValProGluLeuLysPro 639

RESULT 7
D96730
unknown protein F5A18.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96730
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frager, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96730
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-884 <STO>
A:Cross-references: GB:AE005173; NID:G6453896; PIDN:AAF09079.1; GSPDB:GN00141
C:Genetics:
A:Gene: F5A18.20
A:Map position: 1

Alignment Scores:
Pred. No.: 0.000399 Length: 884
Score: 171.50 Matches: 129
Percent Similarity: 34.48% Conservative: 81
Best Local Similarity: 21.18% Mismatches: 208
Query Match: 2.38% Indels: 191
DB: 2 Gaps: 29

US-09-302-812-3 (1-4069) x D96730 (1-884)
QY      83 GCAGCGCAGTCCATCTCTCAGGTTAGTGAATCAGGCTCTCGCGGGCGCGCCCG 142
Db      287 GlyGlyProValHisThrGluGln-----HisMetGlnIrrAlaTyrgly----- 301
QY      143 GCACAGTGGCTGCTGGTCCCGCAGC-----ATGAATCGGGCGCGCCCGCTGTGAA 190
Db      302 AspGlnSerAlaAlaProProSerAsnPheSerAspHisAla----- 316
QY      191 CCTGTCCACCAAGCGACCGCTGGGGCGCGCTACAACTTCG-----CCG 235
Db      317 -----TrrGlnProHisThrThrSerGlyValValTyrrProPro 329
QY      236 GCTGCTTCGGACCGCGGAGCTTTCGAGCAGCAGCAGCGCGGTC---CTCGACCGCCCAAG 292
Db      330 IleProSerSerAlaGlnSerIleProGlnHisAspSerSerMetAlaIleProProVal 349
QY      293 GAGGCTCACGTGCAG-----TTGAGGTCCTCCACCGCTTCGCGAGCTGCGTC--- 340
Db      350 SerGlyHisIleMetProProTyrrGlyArgPheProProAsnProGlnProValGly 369
QY      341 ---CCAGGGCAGCGCGGA-----CAGCAC 361
Db      370 ProProTyrrAlaPheGlyThrLysProProLeuHisProValAlaAlaPheMetAspAsp 389
QY      362 AGAGGAGCGCCACCTCGCTGTGTTTTCACACAAAGACTATTACCAAGTTGGATGGACACT 421

```

```

Db      390 SerTyrrAlaAlaSerSerValProProLysLysAlaProValProAsnTripleLysGlu 409
QY      422 AAAGGAATCAAGACA-----GCCGAATCA 445
Db      410 GluLeuLeuLysLysLysAlaAspLeuGlyArgProSerSerGlyArgPheGluGluArg 429
QY      446 GAAAGTTTGGAT----- 457
Db      430 GluSerMetAspAspValLeuTyrrLysProProThrLysAlaAspGlnProAspLys 449
QY      458 -----ACTAAAGAAACAAAC----- 472
Db      450 LysSerPheSerProSerAsnSerSerAspGluGluGluAspGluMetAspAlaAla 469
QY      473 -----AATACAGATCAATCCATGATGAGTTCGTGA-----CAA 508
Db      470 ArgThrThrGluIleAsnMetGluIleLysArgIleLeuThrGluValLeuLeuLysVal 489
QY      509 AAAGATACCTTTTACCAACATAATGTAGAAAATTTAGTAAATTTCTTCAG---CTAAGT 565
Db      490 ThrAspGluLeuPheAspGluIleAlaThrLysValIleAsnGluAspGluAlaIlePro 509
QY      566 CTTCATAGTCACTCACTCACTGAAAAA-----AGTACACAGTATTGTAACCG----- 610
Db      510 LysAspAspSerValGlnHisAsnHisLysLeuSerSerSerLeuLeuSerThrAlaAsp 529
QY      611 -----CATGACACTGCGAGCATGTGTAGTGGCAAAATGAGGAGAAACACACGAGCAG 664
Db      530 ProLeuHisLysAlaSerAlaLysIleLeuValSerValGluGlyAlaAsnThrLysAla 549
QY      665 CTTTGGAAAGTGAACCTCAACACAGTAACTCCCTGTACCGAGCAGTTTAGTAACTTAAC 724
Db      550 SerSerGlySerProAlaAspValLeuGlyLeuAla-----SerTyrrAlaSer 565
QY      725 ATTGATCGGTCACTCAAAATGATGATCAGTGCAGACAGATGATGAGAGAGAT----- 778
Db      566 AspAspAspAlaAspThrAspAlaAlaSerAspAlaAsnAlaAspGluAsnGlyVal 585
QY      779 -----AGACACAT-----CAACAGTTTCTCACTCACTGTAAGCTT 814
Db      586 GluSerLeuGlyValGlySerArgHisAsnValSerGlnGlnProSerThrGluLysLeu 605
QY      815 GCAATGCAAGCAGACTACGAGATGAACACCGCAGAGAGCAAAAGCCACCCAGAG 874
Db      606 -----ProAspProGluAlaMetAlaSerAlaLysLeuAspProAla 619
QY      875 TGCACAGACTTTCCTCCCTCGGGAAGACTGTGCAAGTTGTGCAGCAAGAT----- 925
Db      620 ValGlyValAsnAlaAsnSerGlyLysAsnSerLysSerGlyLeuGluAspTyrrSerGln 639
QY      926 -----CAGATAGACGTGGTGCCAAAGAGTCCATTTGTTCAGATGT--- 964
Db      640 MetProGlySerThrArgLysAspAspGluAlaGlySerThrLysIleSerAspValSer 659
QY      965 -----GGCTCTGAGGATGTGTGCTACTGGGTCAAAAATGACACAAATTTGATTAGA 1015
Db      660 AlaSerSerGlyLeuAspAspAspThrSerGlySerArgLysGluHis----- 675
QY      1016 CAAGAAAGTTGCTAGGAAATTTCTCTCCATTTGAGAAAGAAAGTGAACCCGATCACCG 1075
Db      676 -----ProAspArgThrAspSerAspLysAspAlaIle 686
QY      1076 ATGGATGTGATTAATCTTAAATAATAGTTCTCCCAACAGCAATAAATCAAGCTCAAG 1135
Db      687 LeuAspGluProHisValLysAsnSer----- 695
QY      1136 CCAGGTTTGTGATGAACAGAGATGTAGTTCTCCCAACAGCAATAAATCAAGCTCAAG 1195
Db      696 -----GlyValLysSerAspCysAsnLeuArgGlnAspSerAsnLysProTyrrGly 712
QY      1196 TTCCAGCAAGAGAGCTGACATTTAGTAAACCG---TACTCTACTAGGCGGT 1252
Db      713 LysAspLeuSerAspGluValSerThrAspArgSerArgIleValGluThrLysGlyGly 732

```



Db 1431 LysSerAsnIleSerMet-----AspGluValPheAsnGluSerGlnAsnGlyGlnLys 1448  
 QY 1385 GATTCTTAAATCAGACATCATTTGATGAGACTGCCAAAGCAGAGCAGACAGAAAGAA 1444  
 Db 1449 AspSerSerAsnIleAsp----- 1454  
 QY 1445 CAGTGGGAAACCAACATCAAGAAACAGAAAGAGATCCCTTAATACGTTCCACCTCAC 1504  
 Db 1455 MetLysGluThrAspMetProGluLysGluArgAspGlnArgTyrValAspValHis 1474  
 QY 1505 CTTTCTCCAGATAAGAGTGGTGGAACTCCCAATGAGGATGAGAGAAATGCTCGG 1564  
 Db 1475 -----ArgAspLysLys-----ProPheGluAsnGlyGluPheGluProThr 1488  
 QY 1565 TGTGGATCCGGCTCCTCTTGAGACCATCTGCCAATCACACAGTAACTATTCGGTGA 1624  
 Db 1489 PheAsnGlySerLysIleSerAsnGluProLysGlnIleSerIleThrThrIleAsnLeu 1508  
 QY 1625 GATCTTTTGGCAGCAGAGAGAGTCTCTAAACCTTTTCCCAACATATAAGATTGTGG 1684  
 Db 1509 Asp-----AsnValPheProThrGluGluProLysLeuVal 1520  
 QY 1685 GATACACACATGTTAAATGCTTGTTCAGACAA----- 1720  
 Db 1521 AlaGluAspAsnCysGluIleGluAlaGluGluArgIleArgLysArgIleLysGln 1540  
 QY 1721 -----AATTGTACCCAGTGAAGATGAG 1744  
 Db 1541 PheGluArgThrThrGlyGluGlnGluIleLeuLysAsnSerGluProAlaGluAspGlu 1560  
 QY 1745 AATGTGAG-----CGAACTCGGGGAGCCGGTGGAGCTCATTCAGACTGCA 1792  
 Db 1561 ThrSerAspGluLysLysHisArgThrAlaAlaValSerIleAspLeuAspLysValPhe 1580  
 QY 1793 CTTCTCAACAATTTACACGACCCCAAC-----TTGAAGATGCT 1834  
 Db 1581 ValGlnGlyThrAlaLysLysProGluAsnAspGluPheAspGluLysIleLysArgGly 1600  
 QY 1835 ATTCTGAATAC----- 1849  
 Db 1601 IleAlaGluPheGluArgSerLysGlnGluLysGluValGluArgSerGlyValAlaGlu 1620  
 QY 1850 GTGGCATATCTAAGAAATGGACATTACAGCTTTGATCGATTTCTGGATTAAGTACTT 1909  
 Db 1621 ThrSerHisSerSerLysHisIlePheAspGluSerAsnIleSerMetAspAspValPhe 1640  
 QY 1910 GAAGAAGCAGAGCTCAACATTTATATCATGCTC----- 1942  
 Db 1641 AsnThrSerGlnLys-----TyrLysSerAspGluLysLeuSerThrProGluArg 1657  
 QY 1943 ---ATCTTGCCTGATATGTTGAAAATTTGCACCTGTCTGCCAAATATTTGCCACCCAGCCA 1999  
 Db 1658 ThrValGluProGluValSerThrAlaThrMetAsnLeuAspAsnIlePheAlaSer 1677  
 QY 2000 ATACACTCCTGAAACAGAGATGAATCATTCATCACAATGTCGACGACAGATGTC 2059  
 Db 1678 GlyIleAlaThrArgGluLys-----AsnThrAspValLeuGluGluArgIleGln 1696  
 QY 2060 AGTCITTTAGCTAAATGCTTTCTCTGCACTTTTCCACGACGA----- 2101  
 Db 1697 LysArgValGluGlu-----PheLysLysThrThrGluAsnLeuGluIle 1711  
 QY 2102 -----AATGCTAAGATGAATCGGAGTAT 2125  
 Db 1712 GlnLysGluValValLeuThrLysGluGluGlyAspAsnSerAspValLysAspHisLys 1731  
 QY 2126 TCTAGTATCCAGACATTAATCTCAATCGATTTGTTGGAGGAGCTTCATCAGAGAAACCG 2185  
 Db 1732 AlaSerAlaValAsnIleAspLeuAspAspValPheIleGlnArgSerSerLysHisPro 1751  
 QY 2186 -----GGAACCTTAAACCGCTCTTCTGCTACTTTAGACAGTCA 2227  
 Db 1752 GluAsnAspGluAspAspGluLysIleArgArgGlyIleAlaGluPheGluArgThrLys 1771

QY 2228 GAGAAAAAACCTACT-----GGTGTGCTGACATTTTACAAGACAGAGTCTTGAAGAT 2278  
 Db 1772 GlnGluLysGluAlaGlnArgSerThrValIleGluThrGlnTyrSerSerLysAsp 1790  
 RESULT 9  
 T41581  
 hypothetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 11-Aug-2003  
 C:Accession: T41581  
 R:Murphy, L.; Harris, D.; Wood, V.; Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z22002  
 A:Accession: T41581  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4717 <MUR>  
 A:Cross-references: EMBL:AL031546; PIDN:CAA20864.1; GSPDB:GN00068; SPDB:SPCC737.08  
 A:Experimental source: strain 972h(-)  
 C:Genetics:  
 A:Gene: SPDB:SPCC737.08  
 A:Map position: 3  
 C:Superfamily: midasin (AAA ATPase with von Willebrand factor type A (vwa) domain)  
 Alignment Scores:  
 Pred. No.: 0.00124 Length: 4717  
 Score: 167.00 Matches: 179  
 Percent Similarity: 34.63% Conservative: 123  
 Best Local Similarity: 20.53% Mismatches: 301  
 Query Match: 2.32% Indels: 269  
 DB: 2 Gaps: 46  
 US-09-302-812-3 (1-4069) x T41581 (1-4717)

QY 434 ACAGCGGATCAGAAAGTTTGGATAGTAAGAAACAACAATACAAGATAGATCCATG 493  
 Db 3942 ThrAlaAsnGlnSerAspLeuAsp-----GluSerGluAlaArgGluLeuGluSerAsp 3959  
 QY 494 ATGAGCTCTGTCAAAAAGATAAATCTTTTACCACCAATAAT----- 532  
 Db 3960 MetAsnGlyValThrLysAspSerValValSerGluAsnGluAsnSerAspSerGluGlu 3979  
 QY 533 -----GTAGAAAAATTAGTAATAGTTTCTCAGCTAAGCTTGTGATAGTCACTC 580  
 Db 3980 GluAsnGlnAspLeuAspGluValAlaAsnAspIleProGluAspLeuSerAsnSerLeu 3999  
 QY 581 ACTGAAAAAAGTACACAGTATTTGAACCCAGCATCAGATCGACCAATGTGTAAAGTGGCAA 640  
 Db 4000 AsnGluLys-----LeuTyrAsp 4005  
 QY 641 AATGAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGATACCCCTGGTA 700  
 Db 4006 GluProAsnGlu-----GluAspLeuLeuGluThrGluGlnLysSer-----Asn 4020  
 QY 701 CCAGAGCAGTTTGTAGTAATGCTTAACATTCATCGTCACTCAAAATGATGATCAC----- 754  
 Db 4021 GluGlnSerAlaAlaAsnAsnGluSerAspLeuValSerLysLeuAspAspAsnLysAla 4040  
 QY 755 ---AGTGACACAGATAGTAGAGAAATAGACAAATCAACAGTCTTCTCACACTGTAAAG 811  
 Db 4041 LeuGluAspLysAspArgGlnGlnLysGluAspGluGluMetSerAspAspValGly 4060  
 QY 812 CTTGCAATTCGAAACAGACACTACGGAAGATGAACCGCAGAGAGCCCAAGCCACCG 871  
 Db 4061 Ile-----AspAspGluIleGlnProAspIleGlnAsnAsn 4073  
 QY 872 AAGTCGACGAAGCTTGTCCATCCTCGGGGAAGTGTGCAAGTGTGCACCAAGAT----- 925  
 Db 4074 -----SerGlnProProGluAsnGluAspHisLeuAspLeuProGluAspLeuLys 4091  
 QY 926 -----GAGATAGCTGTGTGCCAAAGAGTCCATTCTCAGATCTTGCTGCTGAG 973

Db 4092 LeuAspGluLysGluGlyAspValSerLysAspSerAspLeuGluAspMetAspMetGlu 4111  
 QY 974 GATGTTGGTACTGGGTCAAAATAATGACAAATAATGATTAGACAGAAAGATTCCCTAGGA 1033  
 Db 4112 -----AlaAlaAspGluAsnLys-----GluGluAlaAspAlaGlu 4123  
 QY 1034 AATTCCTCTCCA-----TTTGAGAGGAAGTGAACCCGAATCACCAGTGGATGG 1084  
 Db 4124 LysAspGluProMetGlnAspPheGluAspProLeuGluGluGluAsnThrLeuAspGlu 4143  
 QY 1085 GATAATCTCTAAATAGTTCTCAAGACTCAGACGACGATGAGGAG-----ACAAGTCCA 1138  
 Db 4144 AspIleGlnAspAspPheSerAspLeuAlaGluAspAspGluLysMetAsnGluAsp 4163  
 QY 1139 GGTGTTTATGAACAA-----GAAGATGGTAGTTCTCTCCCAA 1174  
 Db 4164 GlyPheGluGluAsnValGlnGluAsnGluSerThrGluAspGlyValLysSerAsp 4183  
 QY 1175 ACAGCA-----AATAAACCCT----- 1189  
 Db 4184 GluGluLeuGluGlnGlyGluValProGluAspGlnAlaIleAspAsnHisProLysMet 4203  
 QY 1190 -----TCAGGTTCCAGCAAGACGACGCTGACATTGCAATTTAGGAAACGCTACTCT 1240  
 Db 4204 AspAlaLysSerThrPheAlaSerAlaGluAlaAspGlu-----Asn 4218  
 QY 1241 ACTAAGGCGCGTGAAGTTAGATTACATTTCCTCAATTT-----GAAGGAGGAGAGATCGC 1294  
 Db 4219 ThrAspLysGlyIleValGlyGluAsnGluGluLeuGlyGluGluAspGlyAlaAlaGlu 4238  
 QY 1295 ACTGGAATGAATGATTTAAATGCTAACTACTCGGAATATTTCTAGCCCTGAATGTAGAA 1354  
 Db 4239 SerGlyValArgGly---AsnGlyThrAlaAspGlyGluPheSerSerAlaGluGlnVal 4257  
 QY 1355 TGAGAAATCTAAGCAACATGGAAGAAAGATTTCT----- 1390  
 Db 4258 GlnLysGlyGluAspThrSerThrProLysGluAlaMetSerGluAlaAspArgGlnTyr 4277  
 QY 1391 ---AAAATCACAGATCATTTGATGAGACTGCCAAAGCAGAGGACAGACAAAGAAAGAACAG 1447  
 Db 4278 GlnSerLeuGlyAspHisLeuArgGluTyrGlnGlnAla-----AsnArgIleHisGlu 4295  
 QY 1448 TGG-----GAAACCAACATCAAGAACAGAAAGAAAGATCCCTAAATACGTT 1495  
 Db 4296 TrpGluAspLeuThrGluSerGlnSerGlnAlaPheAspAspSerGluPheMetHisVal 4315  
 QY 1496 CCACCTCACCTTTCTCCAGATAAGATGGCTTGAACT-----CCCATTTGAG 1543  
 Db 4316 LysGluAspGluGluGluAspLeuGlnAlaLeuGlyAsnAlaGluLysAspGlnIleLys 4335  
 QY 1544 GAGATGAGAAAGATGCTCGGTGGGATCGGCTGCTCTCTTTGAGACCAATCTGCCAAT 1603  
 Db 4336 SerIleAspArgAspGluSerAlaAsn-----GlnAsnProAspSerMet 4350  
 QY 1604 CACACGTAATACTATCGGTAGATCTTTGCGAGCAGAGAGATTCCTAAACCTTTTCCA 1663  
 Db 4351 AsnSerThrAsnIleAlaGluAsp-----GluAlaAspGluValGlyAsp----- 4365  
 QY 1664 ACATTTATAAGATTTGTGGATAACAAGCATGTTAAATGCTTGTTCAGAACAAAT 1723  
 Db 4366 -----LysGlnLeuGlnAspGly-----GlnAsp 4373  
 QY 1724 TTGTACCCAGTGGAGATGAGATGGTGAGCAACTCGCGGGAGCCGGTGGGAGCTCAT 1783  
 Db 4374 IleSerAspIleLys---GlnThrGlyGluAspThrLeuProThrGluPheGlySerIle 4392  
 QY 1784 ---CAGACTGCACCTCTCAACAAATTTACAGCCCCCAAACTTTGAAGGATGCTATTCTG 1840  
 Db 4393 AsnGlnSerGluLysValPheGluLeuSerGluAspGluAspIleGluAspGluLeuPro 4412  
 QY 1841 AAATACATGTGGCATATTTCTAAGAAATGGGACTTTACAGCTTTTGATC----- 1888  
 Db 4413 AspTyrAsnVal-----LysIleThrAsnLeuProAlaAlaMetProIleAspGlu 4429

QY 1889 -----GATTCCTGGATTAAGTACTTGAAGAAGCAGAGCTCAACATTTATATCATGATCC 1942  
 Db 4430 AlaArgAspLeuTrpAsnLys---HisGluAspSerThrLysGln----- 4443  
 QY 1943 ATCTTCCTGATATGTTGTAATAATGCACCTCTGCTCTGCCAAATATTTGCAACCCAGCCAATA 2002  
 Db 4444 -----LeuSerIleGluLeuCys----- 4449  
 QY 2003 CCACCTCTGAAACAGAAAGATGAATCATTCATCACAATGTCGACGAGAACAGATTGCCAGT 2062  
 Db 4450 -----GluGlnLeuArgLeu 4454  
 QY 2063 CTTTTRAGCTAATGCTTCTTCGACATTTCCACGACGAAATGCTAAGATGAATCGGAG 2122  
 Db 4455 IleLeuGlu-----ProThrLeuAlaThrLysMetGlnGlyAsp 4467  
 QY 2123 TATTCCTAGTTACCCAGACATTAACCTTCAATCGATTCTTTGAGGGACGTTCAATCAAGG--- 2179  
 Db 4468 PheArgThrGlyLysArgLeuAsnMetLysArgIleIleProTyrIleAlaSerGlnPhe 4487  
 QY 2180 AAACCCGGAACAACTTAAACCGCTCTTCTGCTACTTTTAGAAGAGTCAACAGAAAAAACCT 2239  
 Db 4488 LysLysAspLysIle-----TrpMetArgArgValLysProSerLysArg 4502  
 QY 2240 ACTGGGTTGGTGACATTTACAGACAGAGTCTTGAAGATTTCCAGAAATCGGGAAGATGT 2299  
 Db 4503 ThrTyrGlnValMetIle-----SerIleAspAsp----- 4512  
 QY 2300 GAAAAACCCCTTGACACGATTGCATGTCACCTTACGAAGGTACCATAGAAAAATGGCAA 2359  
 Db 4513 -----SerLysSerMetSerMetSerGluSerGlySer 4521  
 QY 2360 GGCATGCTACAGTGGATTTTGCAATCGTTTGTGTGGAGGTGGTAAACCATGCGAGGA 2419  
 Db 4522 ThrValLeuAlaLeuGluThrLeu-----AlaLeuValThrLysAlaLeu 4536  
 QY 2420 CTTGTGCAGAA-----GAAATCCGCTTTTAAATCAATCCTGAGTTGATT 2464  
 Db 4537 SerLeuLeuGluValGlyGlnIleAlaValMetLysPheGlyGluGlnProGluLeuLeu 4556  
 QY 2465 -----ATTTCACGGCTCTTCACTGAGGTGCTGATCACAATGAATGCTTAATTATCACA 2518  
 Db 4557 HisProPheAspLysGlnPheSerSerGlu-----Ser 4567  
 QY 2519 GGTACTGACAGCTACGATGATGAATACAC----- 2545  
 Db 4568 GlyValGlnMetPheSerHisPheThrPheGluGlnSerAsnThrAsnValLeuAlaLeu 4587  
 QY 2546 -----GGCTATGCTGAGACATATCGTTGGTCCCGGAGCCAC 2581  
 Db 4588 AlaAspAlaSerMetLysCysPheAsnTyrAlaAsnThrAlaSerHisHisArgSerAsn 4607  
 QY 2582 GAAGATGGGAGTGAAGGGAGCGACTCGAGCGCGCTGCACTGAGATCGTTGCCATCGAT 2641  
 Db 4608 SerAsp-----IleArgGlnLeuGluIleIleSerAsp 4619  
 QY 2642 GCTCTTCACTTACAGCGCTACCTCGATCGATCAGTTGTGCTGAGAAATGAGACCGGAGCTG 2701  
 Db 4620 GlyIle-CysGluAspHisAspSerIleArgLysLeuLeuArgArg-----AlaGlnGln 4637  
 QY 2702 AACAGGCTTACTGATTTCTCGTCTCTGGAG 2735  
 Db 4637 uGluLysValMetIleValPheValIleLeuAsp 4648

## RESULT 10

I38614

helicaase II - human

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999

C;Accession: I38614

R;Stacyton, C.L.; Dabovic, B.; Gulisano, M.; Geetz, J.; Broccoli, V.; Giovanazzi, S.; Boss

Hum. Mol. Genet. 3, 1957-1964, 1994

A:Title: Cloning and characterization of a new human Xq13 gene, encoding a putative heli

A:Reference number: 138614; MUID:95179111; PMID:7874112

A:Accession: 138614

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1641 <RES>

A:Cross-references: EMBL:U09820; NID:g606832; PIDN:AAC50069.1; PID:g606833

C:Genetics:

A:Gene: RAD54L

Alignment Scores:  
 Pred. No.: 0.00105 Length: 1641  
 Score: 166.50 Matches: 178  
 Percent Similarity: 36.71% Conservative: 134  
 Best Local Similarity: 20.94% Mismatches: 322  
 Query Match: 2.31% Indels: 216  
 DB: 2 Gaps: 48

US-09-302-812-3 (1-4069) x 138614 (1-1641)

389 AAACAAAGACTATTACAGTTGGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAA 448  
 277 GluArgArgAsnLeuSerSerLysArgAsnThrLysGluLeuGlnSerGlySerSer 296  
 449 AGTTTGGATAGTAAGAAAC-----AACATACAAGATAGATCCATGATGATCT 502  
 297 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysGlnArgThrSer 315  
 503 GTACAAAAGATAAATTTTACCAATATGTAGAAAATAGTAAATGTTTCTCAGCTA 562  
 316 LysLysLysAlaValIleValLysLysLysArgAsnSerLeuArgThrSerLys 335  
 563 AGTCTTGATAGTCACTACTGAAAGAAACACACGGAGACCTTTTGGAAAGTCAACT 682  
 336 ArgLysGlnAlaAspIleThrSerSerSerSerAspIleGluAspAspGlnAsn 355  
 623 GCAATGTGTAGTGGCAAAATGAAGGGAACACACGGAGACCTTTTGGAAAGTCAACT 682  
 356 SerIle-----GlyGluGlySerSerAspGluGlnLysIle-----LysPro 369  
 683 CAACAGTAACCCCTGGTACCAGAGCAG-----TTTAGTAATGCTAACATTCATCGG 733  
 370 ValThrGluAsnLeuValLeuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 389  
 734 TCA-----CCTCAAAATGATGATCAGTGCAGTGCACACAGATAGTGA 772  
 390 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspProGlu 409  
 773 GAGAAATAGACAAATCAACAGTTTCTCAACTGTAAAG-----CTTGCAATGCAAG 826  
 410 AsnArgIleAlaLysLysMetLeuLeuGluGluLysAlaAsnLeuSerSerAspGlu 429  
 827 CAGACTACGGAAGATGAACACCCAGAGAGAGCCAAAGCCACAGAGTGCAGCAAGTCT 886  
 430 AspGlySerSerAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 449  
 887 TCCCATCTCTGGGAAGACTGTCTCAGTGTCTCAGCAAGATGAGATAGACGTGGTGCMAAG 946  
 450 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer----- 463  
 947 AGTCCATGTCTCAGATGTTGGCTCTCAGGATGTTGGTACTGGCTCAAAAATGAC----- 1000  
 464 -----GluSerAspSerAspSerGluGlu-----SerLysLysProArgTyr 477  
 1001 ---AACAAATGATTAGACAGAAAGTTGCTCCTAGGAAATTCCTCCCA---TTTGAGAG 1054  
 478 ArgHisArgLeuLeuArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLys 497  
 1055 GAAAGTGAACCCGATCACCAGTGGATGTTG-----GATAATCTTAAATAGTCT 1105  
 498 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 517  
 1106 CAAGACTCAGAGCAGATGAGGAGACAGTCCAGGTTTGTGATCAACAAAGATGGTAGT 1165

518 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal-----Ser 535  
 1166 TCCTCCCAACAGCAAAATAAACCTTCAAGTTCCAAAGCAGAGACGCTGACATTAATTT 1225  
 536 GluSerGluAspGluGlnArgProArgThrArgSerAlaLysLysAlaGluLeuGluGlu 555  
 1226 AGGAAACGG---TACTCTACTAAGCGCGTGAAGTGTAGATACATTTTCCATTTCAAGGA 1282  
 556 AsnGlnArgSerTyrLysGlnLysLysLysArgArgIleLysValGlnGluAspSer 575  
 1283 GGA----- 1285  
 576 SerSerGluAsnLysSerAsnSerGluGluGluGluGluLysGluGluGluGluGlu 595  
 1286 -----GAGAGTCGCACTGGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1327  
 596 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 614  
 1328 GGAATATTTCTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1381  
 615 Gly-----LysGlyArgLysLysLysLysLysLysLysLysLysLysLysLys 625  
 1382 AAGGATTTCTAAATC---ACAGATCATTTGATGACTGCCCAAAAGCAGAGACAGAGA 1438  
 626 LysAspAspLysLeuArgThrGlnAsnAlaLeuLysGluGluGluGluGluGluGlu 645  
 1439 AAAGNACAGTGGGAAACCAAAACATCAAGAACAGAGAGATCCCTAAATACCTTCCA 1498  
 646 LysArgIleAlaGluArgGluGluGluGluGluGluGluGluGluGluGluGluGlu 663  
 1499 ---CCTCACCTTTCTCCAGATAAGAGTGGTGGAACTCCCAT----- 1540  
 664 IleGluAspAlaSerProThrLysCysProIleThrLysLeuValLeuAspGluAsp 683  
 1541 GAGGAGATGAGAGAAATGCTCGGTGGGATCGGCTCGCTCTCTTGAGACCATCTGCC 1600  
 684 GluGluThrLysGlu-----ProLeuValGlnValHisArg 695  
 1601 AATCACAGTAACCTATTTCCGGGTAGATCTTTTGGAGCAGGAGAGTTCCTTAACCTTT 1660  
 696 AsnMetValIleLysLeu-----LysPro----- 703  
 1661 CCAACACATTTATAAGAT-----TTGTGGATAACAAGCATGTAAAAATGCT 1708  
 704 -----HisGlnValAspGlyValGlnPheMetTrpAsp----- 714  
 1709 TGTTCAGAACAAATTTGTACCCAGTGAAGATGAGATGGTGAAGACCTCGGGGAGC 1768  
 715 CysCysCysGluSerVal-----LysLysThrLysLysSerProGlySer 729  
 1769 CGTGGGAGCTCATTCAGACTGCATTTCTCAACAAATTTACACGCCCAAACTTGAAG 1828  
 730 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys-----ThrLeuGln 744  
 1829 GATGCTATTTGAAATACAAATGGCATATTTCTAAGAAATGGACTTT---ACAGCTTTG 1885  
 745 ValValSerPheLeuHisThrValLeuLeuCysAspLysLeuAspPheSerThrAlaLeu 764  
 1886 ATCGATTTCTGGATTAAGTACTT-----GAAGAAGCAGAGCTCAACAT 1930  
 765 ValGlyLeuSerSerSerIleLeuAlaPheAsnTrpMetAsnGluPheGluLys----- 782  
 1931 TTATATCAGTCCATCTGCTGATATGTTGAAATTTGCATCTGTCTGCCAAATATTTGC 1990  
 783 ---TrpGlnGluGlyLeuLysAspGluLysLeuGlu-----ValSerGluLeuAla 799  
 1991 ACCCAGCAATACCACTCTCTGAAACAGAGAGATGATCAATCATCATCAATGTGCGAGAA 2050  
 800 ThrValLysArgPro-----GlnGluArgSerTyrMetLeuGlnArgTrpGlnGlu 816  
 2051 CAGATTGCGAGTCTTTTACGTAATGCTTTCTTCTGCACTTTCCACGAGAAATCCTAAG 2110





Db 1058 GluThrValGlnCysAsnLysSerSerCysGluGluThrSerIleThrLysValAlaLys 1077  
 QY 1202 GCAAGAGCGTGACATTGAATTTAGGAAACGGTACTTACTTAAGGCGGTGAA----- 1255  
 Db 1078 AsnIleGluProHisValIlePheArg-----SerThrGluAlaSerAspMetAla 1094  
 QY 1256 GTTAGATTACATTTCCAAATTTGAAGGAGGAGAGTCGCACATGGAATGAATGAATTAAT 1315  
 Db 1095 ValGlyIleHisTyrThrLeuArgSerSerGluLysValGluGluThrGluGluIleLys 1114  
 QY 1316 GCTAAACCTACCTGGAATATTCTAGCCCTGAATGTAGAATGCAGAAATTTCTAAGCAACAT 1375  
 Db 1115 ---AsnIleAlaArgAsnGlyGlySerAlaThrPheSerCysPheAlaSerGlyAsp--- 1132  
 QY 1376 GGAAGAAAGGATCTTAAATCATCAGATCATTTGATGAGACTGCCCAAGCAGAGGACAGA 1435  
 Db 1133 ---GluSerProAspSerValSerAlaPheLeuThrArgGlnProGlnGluGlu----- 1149  
 QY 1436 AGAAAGAACAGTGGGAAACCAACATCAAGAACAGAGGAGAGTCCCTAAATACGTT 1495  
 Db 1150 -----ThrThrGlu-----LysLeuPhe 1155  
 QY 1496 CCACCTCACCTTTCCAGATAAGAGTGGCTTGGAACTCCCATTCAGGAGATGAGAAGA 1555  
 Db 1156 ProThrProMetPheAspPheIleLysPheAsnSerThrAlaAlaGluGluPheAlaVal 1175  
 QY 1556 ATGCCTCGTGTGGATCCGGCTGCTCTCTTGAGACCATCTGCCAATCACAGTAACCT 1615  
 Db 1176 Trp-----AsnThrThrIlePhe 1181  
 QY 1616 ATTCGGGTAGATCTTTTGGAGCAGGAGAGTTCCTPAAACCTTTTCCACACATTAATAA 1675  
 Db 1182 ArgArgLysAsp-----AsnGluGlyGluValGluLysIlePheAsnThrSer----- 1197  
 QY 1676 GATTTGTGGGATAACAAGCATGTAAATGSCCTTGTTCAGAACAAATTTGTACCCAGTG 1735  
 Db 1197 ----- 1197  
 QY 1736 GAAGATGAGATGTGTGAGCGAACTGGGGGAGCGGTGGAGCTCATTCAGACTGCACCTT 1795  
 Db 1198 GluAlaGlyHisAsnGluThrPheSerAlaAsnAlaAlaGluAspValSerValThrLeu 1217  
 QY 1796 CTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAAATACAATGTCGCA 1855  
 Db 1218 -----AspAlaAspLeuHisPheGlyValGly 1226  
 QY 1856 TATCTTAAGAAATGGGACTTTTACAGCTTTGATCGATTTCTGGGATAAGGTA----- 1906  
 Db 1227 TyrLysGluHisArgGlnIleThrLys-----AspGluAlaAsnGlnGly 1241  
 QY 1907 -----CTTGAAGACGAGNAGCTCAACATTTATATATACAGTCCATCTTGCTGAT 1954  
 Db 1242 GluGlyThrGlyMetHisSerGlyAlaSerGluGluThrIlePheAsnLeuGlyTyrAsp 1261  
 QY 1955 ATGCTGAAATTCACATCTGTCTGCCAAATATTGCAACCCAGCCCAATACCATCTCTGAA 2014  
 Db 1262 TyrCysLysGlnProThrGluPheThrValCysValThrGluAspLysLeuLeuIle 1281  
 QY 2015 CAGAGATGAATCATTCATCAATGTCCGAGGAACAGATGCCCAGTCTTTTAGCTAAT 2074  
 Db 1282 GlnGlyAla---TyrGlyPheArgAlaAlaLysGluGluSerIleThrLeuAspAlaAsp 1300  
 QY 2075 GCTTTCTCTGCACATTTCCAGACCAATGCTTAAGATGAATCGGAGTATTCTAGTTAC 2134  
 Db 1301 LeuHisPheGlyValAsnTyrArgAspLeuLeuAlaMetGlyLeuHisAlaSer----- 1318  
 QY 2135 CCAGACATTAACCTCAATCGATTGTTGAGGGACGTTTCATCAAGGAAACCGGAGAACTT 2194  
 Db 1319 -----AsnAsnGluIleGluGlyThrGlyMetArgSerThrAlaSerGlu 1333  
 QY 2195 AAAACGCTCTTC-----TGCTACTTTAGAGAGTCTCAGAGAAAAA 2236  
 Db 1334 GluThrIlePheAsnLeuAlaTyrAspTyrCys-----LysGln 1346

QY 2237 CCTACTGGGTGGTGCATTTTCAAGACAGAGCTCTTGAAGATTTTCCAGAAATGGGAAAGA 2296  
 Db 1347 ProThrGluPheArgThrValPheValSerGluAspHisGlnPheValHisGlyAlaPhe 1366  
 QY 2297 TGTGAAACCCCTTCACACGATTCATGCTTACGAAAGGTACCATCAGAGAAATATGGC 2356  
 Db 1367 GlyPheArgAlaValGlyGluGluHisIle-----GluThrGlnValLeuGlu----- 1382  
 QY 2357 CAAGCATGCTACAGGTGATTTTGCAAATCGTTTGTGGAGGTGGTGTACCACTGCA 2416  
 Db 1383 -----LeuGlnAlaArgMetValGluValMetValGluGlySerValHisLeu 1399  
 QY 2417 GCATTTGTGCAAGAGAAATCCGCTTTTAAATCAATCCTGAGTGTGATTTTACGGCTC 2476  
 Db 1400 AlaArgArgHisGluAspGlnProPheVal-----Leu 1410  
 QY 2477 TTCACTGAGGTGCTGCATCACAAATGAATGCTCTAATATCACAGGTACTGAGCAGTACAGT 2536  
 Db 1411 TyrThrGluValIleGluGluThr-----IleIleArgValAspGluGlnLeuGlu 1427  
 QY 2537 GAATACACAGGCTATGCTGAGACATATCGTTGCTCC-----CGAGCCACGAGATGGG 2590  
 Db 1428 LysLysThrThrValIleGluThrGluGlnAlaSerGluValLysMetArgGluLysGly 1447  
 QY 2591 AGTGAAGGAGCAGCTGCGAGCGCGCTGCACCT-----GAGATCGTTGCC 2635  
 Db 1448 GluGluArgArgLysGluGluLysArgValSerPheAlaAlaGluValGlnGluLysThr 1467  
 QY 2636 ATCGATGCTCTCACTTCAGACGCTACCTCGATCATGTTTGTGCTC----- 2680  
 Db 1468 MetGluAlaIleAspLysSerLeuGlyLeuAspThrSerMetGluValGluProAlaPhe 1487  
 QY 2681 -----GAGAAATGAGACGCGAG 2698  
 Db 1488 GlnLysProSerIleIleLysLysProMetLysLysGluArgGluArgSerArgAsp 1507  
 QY 2699 CTGAACAGGCTTACTGTGGATTTCTCGTCTCTGGAGTTTCTCAGAGAAATCTTTCTGCA 2758  
 Db 1508 LeuArgGlnAsnAlaAlaProAlaPheLysProValArgAsnSerLeuLeuGlnAla 1527  
 QY 2759 GTGGCCACAGGA----- 2770  
 Db 1528 LeuAlaIleGlySerProHisAsnIleProHisPheLysThrLeuAspAspIleValLys 1547  
 QY 2770 ----- 2770  
 Db 1548 AlaIleLysHisAlaGlyLeuGluTyrSerAsnLeuIlePheGlyIleAspTyrThrLys 1567  
 QY 2771 ---AACTGGGGCTGTGGTGCCTTTGGGGTGTATGCCAGG----- 2806  
 Db 1568 SerAsnPheTyrGlnGlyGluArgThrPheAspLysArgProLeuHisThrIleAspPro 1587  
 QY 2807 -----TTAAAGCCTTAATACAGATA-----TTGCAGCTGTCT 2839  
 Db 1588 AlaGluMetAsnProTyrGlnGlnValIleGlnIleValGlyLysThrLeuSerSerPhe 1607  
 QY 2840 GCAGCTGAGCGAGTGTGTTTATTTACCTTTGGGACTCAGAAATGTATGTAGAGACATT 2899  
 Db 1608 AspAlaAspGlyGlnIleProAlaTyrGlyPheGlyAspGluGlu----- 1622  
 QY 2900 TACAGCATGCAC---ATTTC---CTTACTGAAAGGAAACTCACCTGTTGGAGATGTG--- 2950  
 Db 1623 PheThrAspHisGlyIlePheAsnIleAlaGluArgTyrAspLeuGluLysAspCysAsn 1642  
 QY 2951 ---TATAAGCTGTTCTGCTACGATCTACATGAAGAAATGCAGAAATGTTTCCACCTGGA 3007  
 Db 1643 GlyPheGluGluValLeuArgValTyrAsnGluValThrProThrIleGluMetSerGly 1662  
 QY 3008 CCAGACATCAAGCTTATCCATTCATATACCATGCTGCTGAGTCTGCTGTCAGAG 3061  
 Db 1663 Pro---ThrAsnPheValProLeuIleAspArgAlaIleGluIleCysLysGlu 1679

RESULT 12  
T24806  
hypothetical protein T10G3.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24806  
R:Burton, J.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19937  
A:Accession: T24806  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1164 <W1>  
A:Cross-references: EMBL:Z81118; PTDN:CA803330.1; GSPDB:GN00023; CESP:T10G3.5  
A:Experimental source: clone T10G3  
C:Genetics:  
A:Gene: CESP:T10G3.5  
A:Map position: 5  
A:Introns: 15/3; 73/3; 387/3; 412/3; 708/3; 818/3; 894/3; 982/3; 1080/2

Alignment Scores:  
Pred. No.: 0.00285 Length: 1164  
Score: 160.00 Matches: 168  
Percent Similarity: 36.41% Conservative: 171  
Best Local Similarity: 18.05% Mismatches: 360  
Query Match: 2.22% Indels: 232  
DB: 2 Gaps: 33

US-09-302-812-3 (1-4069) x T24806 (1-1164)

QY	374	ACCTCGCTGTTTCAACAAAGAACTATTACCACTGGATGGACACTAAAGGAATCAAG	433
Db	253	ThrGluLeuValAsnAlaGlnLysLeu-----MetAspAla-----	264
QY	434	ACAGCGGAATCAGAAAGTTGGATAGTAAAGAAACAAACAATAGTAATGCCATG	493
Db	265	IleSerGlnGluLysAspIleGluLeuLysGluHisLeuAsnSer---IleArgAsnLeu	283
QY	494	ATGAGTCTCTGACAAAGATAACTTTTACCACATATAGTAAAGAAATAGTA-----	547
Db	284	SerMetGluArgGluLysGlnHisIleValAsnGluAsnLeuGluLysIleGlu	303
QY	548	-----AATGTTCTCAGCTAGCTTGTGATAGTCACTCACTCACTCACTCACTCACT	598
Db	304	GlyGluGluThrValLysGlnLeuGlnIleSerTyrAspAlaGlnSerGluGluLys	323
QY	599	TATTTGAACACCATCAGACTCAGCAATGTGTAGTGGCAAAATGAAGGGAAACACACG	658
Db	324	GlnArgAsnGluArgVal-----GlnLeuGluAlaArgIleGlu	337
QY	659	GAGCAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTGGTACCAGAGCGATTAGTAAT	718
Db	338	GluAsnValPheGlu-----	342
QY	719	GCTAATGATCGGTCACTCAAAATGATGATCAGCTGACAGACAGATAGTGAAGAGAT	778
Db	343	-----LeuSerGlnAsnLysGlnAsnValLysArgLeuGluAspLysValGlnGluSer	360
QY	779	AGAGCAATCAACAGTTTCTCAACTGTAAAGCTTGCAAATGCAAGCAG-----	829
Db	361	GlnAspAlaLeuGlnMetLeuSerAsnIleAsnGlySerAspGluGlnMetIleSer	380
QY	830	-----ACTAGGAAGATGAACACGCCAGACGAAAGCCAAAGCCAC	868
Db	381	LeuAsnSerLysPheGluArgAsnThrAlaGluArgLysArgIleGluAlaValPheGlu	400
QY	869	CAGAACTGACGAAGCTTTGCCATCTCGGGGAGAGCTGTGCAAGTTGTGCAGCAAGATGAG	928
Db	401	GluLysValThrValGln-----GlyGluArgLeuLysThrLeuGluMetAlaAsn	417
QY	929	ATAGACCTGTGCGCAAGAGATCCATGTCAGATGTTGGCTCT-----GAGGATGTT	979
Db	418	LeuAspLeuThr-----AsnGluLeuAlaSerMetGlySerLeuLeuAspLysGluArg	435

QY	980	GGTACTGGTCAAAAAATGACAAATGATTAGTACACAGAAAGTTTCCTAGGAAATTC	1039
Db	436	SerLeuLeuGluGlnLysAsnLysGluIleSerGluArgAspSerSerIleAsnAspLeu	455
QY	1040	CCTCCATTTTGAAGAGGAAAGTGAACCCGATACCCGATCGGATGTGGATAATCTTAA	1099
Db	456	-----LysGluLysLeuAlaGluSerGluLysLysAlaThrLysTyrLysAsn	471
QY	1100	AGTTGTCAACACTCAGACAGCAGATGAGGACAGTCCAGCTTTTGTATCAACAGAGAT	1159
Db	472	Glu-LeuLysGluHisAlaAspLeuValGluAsnLeuThrLeuGlnLeuAsnLysLeu	491
QY	1160	GGTAGTTCTCCCAACAGCAAAATAACCTTCAAGTTCACAGCAGACGCTGACATT	1219
Db	491	n-----GluAsnSerLysAspLeuMetGluLysIle	501
QY	1220	GAATTTAGGAAACGGTACTCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAA	1279
Db	501	eserAlaGlyGluGlyAlaLysMetAla-----IleGluGlnLeuGlu	516
QY	1280	GGAGGAGAGTGCCTGGAATGAATGATTAATGCTAAATCTACCTGGAATATTCT	1339
Db	516	uGlnGluLysValLysLeu---ThrAsnGluLeuGlnThrSerSerGlu-----	531
QY	1340	AGCCTGAATGTAGAATGCAGAAATTTCTAAGCAACATGGAAGAAAGGATCTTAA	1399
Db	532	-----LysThrLysLysAlaSerGlyGluLeuGluAlaLysIleSer	545
QY	1400	GATCATTTGATGAGCTGCCCAAGCAGACAGACAGAAGA-----AAGACAG	1450
Db	546	GluLeuGluLysLysLeuArgAspAlaGluAlaSerArgThrAspLysGlnLysTrp	565
QY	1451	GAACCAACATCAAGACACAGAAAGGAAGTCCCTAAATACGTTCCACCTCACCT	1510
Db	566	LysGlnGluLysGluSerPheGluArgLys-----	575
QY	1511	CCAGATAAGAGTGGCTTGGAACTCCCATTTGAGGAGATGAGAGAAATGCCCTGG	1570
Db	576	-----LeuAlaGluAlaGluAspGluIleLysArgLysGlyGluArgPhe	590
QY	1571	ATCCGGCTGCTCTTGTGAGACCATTCGCAATCACAGTAATATTCGGGTAGATCTT	1630
Db	591	ValGluMetGluLysGluMetGluGluGluArgGlnLysAlaThrAspArgThrLeu	610
QY	1631	TTGCGAGCAGGAGAGTCTCTAAACCTTTCCACACATTAATATAAGATTTTGG	1690
Db	611	Leu-----LysAspAlaLeuValAsn	617
QY	1691	AAGCATGTAAATGCCTTGTTCAGAACAAATTTGTACCCAGTGAAGATGAGATGGT	1750
Db	618	-----SerGluLysAsnLeuGluThrIleLysLysGluSerGlu	630
QY	1751	GAGCGAATCGGGGAGCGGTGGAGCTCATTCAGACTGCACCTCTCTCAACAAATTA	1810
Db	631	AspArgGlu-----LysIleValArgGluLysAspAlaHisLeuGluGlu	645
QY	1811	CGACCCCAAACTTGAAGATGCTATTCTGAAATCAATGGCATATTTCTAAGAAATGG	1870
Db	646	AsnLysLysArgGlyIleGluAspAlaValGlnLysLeuGluAlaGluLysArgAla	665
QY	1871	GACTTTACAGCTTTGATCGATTTCTGGGATAAGTACTTGAAGACGAGACCTCAAC	1930
Db	666	GluLeuGluAlaSerValSerSerArgAspThrThrValSerThrLysGluSerGlu	684
QY	1931	TTATATCATGTCCTTCTGCTGATATGTTGAATTTGCACCTCTGTCTGCCAATATT	1990
Db	685	-----LeuSerGluLeu---LysGlyLysLeuThrGluSerAsnSerPhe	698
QY	1991	ACCCAGCCCAATACCATCTCTGAAACAGAAAGATGAATCATTCATCATGATGCG	2050
Db	699	IleGluGluLeuLysValGlnValGluLysValSerAsnGluIleSerGluLysGln	718



Db 691 IleGlnThrLeuLysGluAspLeuArgLysSerGluGluAlaLeuArgPheSerLysLeu 710  
 QY 854 GAAGCCAAAGCCACAGAGTCAGCAAGTCCTGCGGAGCAAGTCGCAAGT 913  
 Db 711 GluAlaLysAsnLeuArgGluValIleAspAsnLeuLysGlyLysHisGluThrLeuGlu 730  
 QY 914 TGTCAGCAAGATCAGATAGACGTGGTCCCAAGAGTCCTATTCAGATGTCCTCTGAG 973  
 Db 731 AlaGlnArgAsnAspLeu-----HisSerSerLeuSerAsp----- 742  
 QY 974 GATGTTGGTACTGGTCAAAAATGACAACAATGATTAATGATAGACAAGATGTCCTAGGA 1033  
 Db 743 -----AlaLysAsnThrAsnAlaIleLeuSerSerGlu----- 753  
 QY 1034 AATTCCTCCATTTGAGAGGAAGTGAACCGAATCACCGATGATGGGATAATCT 1093  
 Db 754 -----LeuThrLysSerSerGluAspValLysArgLeuThrAla---AsnVal 768  
 QY 1094 AAAAATAGTTGTCAGAGTCAGAGCAGATGAGGAGCAAGTCAGGTTTGTATGACAA 1153  
 Db 769 GluThrLeuThrGlnAspSerLysAla-MetLysGlnSerPheThrSerLeuValAsnSe 788  
 QY 1154 GAAGATGGTAGTTCCT-----CCCAACAGCAATAAACCTTCAAGG 1195  
 Db 788 rTyrGlnSerIleSerAsnLeuTyrHisGluLeuArgAspHisValAsnMetGln-- 807  
 QY 1196 TTCACAGCAGAGCGCTG-----ACATT 1219  
 Db 808 -SerGlnAsnAsnThrLeuLeuGluSerGluSerLysLeuLysThrAspCysGluAsnLe 827  
 QY 1220 GAATTTAGGAAACGCTACTCTACTAAGCGCGTGAAGTTAGATTACATTTCCAAATTGAA 1279  
 Db 827 uThrGlnAsnMetThrLeuIleAspAsnValGln--LysLeuMetHisLysHisVal 846  
 QY 1280 GAGGAGCAGAGTCACATCGA-----ATCAATGATTTAAATGCTAACTACCTGGAAATATT 1336  
 Db 847 AsnGlnGluSerLysValSerGluLeuLysGluValAsnGlyLysLeuSerLeuAspLeu 866  
 QY 1337 -----TCTAGCCTGAATGTAGATGCGAGAAATCTAAGCAACATCGAAAAAAG 1384  
 Db 867 LysAsnLeuArgSerSerLeuAsnValAlaIleSerAspAsnAspGln----- 882  
 QY 1385 GATTTCTAAATACAGATCATTTGATGAGATGCGCCAAAGCAGAGGACAGAGAAGAAAGAA 1444  
 Db 883 -----IleLeuThrGlnLeuAlaGluLeuSerLysAsnTyrAspSerLeuGluGln 899  
 QY 1445 CAGTGGAAACCAACATCAAGAACAGAAAGGAGATCCCTAAATACGTTCCACCTCAC 1504  
 Db 900 -----GluSerAlaGlnLeuAsnSerGlyLeuLys-----Ser 910  
 QY 1505 CTTTCTCCAGATAAGAGTGGCTTGGAACTCCCATTCAGAGATGACAGAGAATGCCTCGG 1564  
 Db 911 LeuGluAlaGluLysGlnLeuLeuHisThrGluAsnGluLeuLeuHis----- 926  
 QY 1565 TGTGGGATCCGGCTCTCTTGTGACCACTCTGCCAATCACAGTAATATTCGGGTA 1624  
 Db 927 -----IleArgLeuAspLysLeu-----ThrGlyLysLeuLysIle 938  
 QY 1625 GATCTTTGCGAGCAGAGAGTTCCTAAACCTTTCCACACATTAATAAGATTGTGG 1684  
 Db 939 GluGluSerLysSerSerAspLeuGlyLysLysLeuThrAlaArgGlnGluGluIleSer 958  
 QY 1685 GATAACAGCATGTATAAATGCTTTGTTCCAGAACAAAATTTGTACCCAGTCGGAAGATGAG 1744  
 Db 959 AsnLeuLysGluLysAsnMetSerGlnSer---GlnAlaIleThrSerValLysSerLys 977  
 QY 1745 AATGGTGAAGCA---ACTGGGGAGCGCGTGGAGCTCATTCAGATGCTCTCTCAC 1801  
 Db 978 LeuAspGluThrLeuSerLysSerSerLysLeuGluAlaAspIleGluHisLeuLysAsn 997  
 QY 1802 AAATTTACAGCACCACAACTTGAGGATGCTATTCTGAAATACAAATGCGCATATTCT 1861

RESULT 14

S65236

probable membrane protein YPL217c - yeast (Saccharomyces cerevisiae)

Db 998 LysValSerGluValGluValGluArgAsnAlaLeuLeuAlaSerAsn----- 1013  
 QY 1862 AAGAAATGGGACATTTTACAGCTTTTGATTCGAT----- 1891  
 Db 1014 -----GluArgLeuMetAspAspLeuLysAsnAsnGlyGluAsnIleAla 1028  
 QY 1892 TTCTGGGATAAGTACTTGAAGACAGAGAGCTCAACATTATATCAGTCCATCTTTCCT 1951  
 Db 1029 SerLeuGlnThrGluIleGluLysLysArgAlaGluAsn----- 1041  
 QY 1952 GATATGGTGAATAATTCACCTCTGCTCGCAAAATATTGCAACCCAGCAGCAATACCATCTCTG 2011  
 Db 1042 AspAspLeuGlnSerLysLeuSerValValSerSerGluTyrGluAsnLeuLeuIle 1061  
 QY 2012 AAACAGAAAGATGAATTCATCACAATGTGCGAGGAACAGATTCGCCAGTCTTTTAGCT 2071  
 Db 1062 SerSerGlnThrAsnLysSerLeuGluAspLysThrAsnGlnLeuLysTyrIleGluLys 1081  
 QY 2072 AATGCT-----TTCTTCGACATTTCCACGACGAAATGCTAAGATGAAATCGGAGTAT 2125  
 Db 1082 AsnValGlnLysLeuLeuAspGluLysAspGlnArgAsnValGluLeuGluLeuThr 1101  
 QY 2126 TCTAGTTACCCACACATTAATTCATCGATTGTTGAGGAGCGTTTCATCAAGGAACCG 2185  
 Db 1102 SerLysTyr-----GlyLysLeuGlyGluGluAsn 1111  
 QY 2186 GAGAACTTAAACCGCTCTCTGCTACTTTAGAGAGTCACAGAGAAAACCTACTGGG 2245  
 Db 1112 AlaGlnIleLysAspGluLeuLeuAlaLeuArgLysLysSerLysGlnHisAspLeu 1131  
 QY 2246 TTGGTCACATTTACAGACAGAGTCCTTGAAGATTTTCCAGATGGGAAAGATCTGAAAAA 2305  
 Db 1132 CysAlaAsnPheValAspAspLeuLysGluLysSerAspAlaLeuGluGluLeuThrAsn 1151  
 QY 2306 CCCTTCACACGATGTGATGTCATCTACGAGGTACCATAGAGAAAATGCCAAGGATG 2365  
 Db 1152 GluLysAsnGluLeuIleValSerLeuGluGlnSer---AsnSerAsnAsnGluAlaLeu 1170  
 QY 2366 CTACAG-----GTGAGATTTTGCAAATCGTTTGTGGAGGTGGTAAACCAGTGCAGGA 2419  
 Db 1171 ValGluGluArgSerAspLeuAlaAsnArg----- 1180  
 QY 2420 CTTGTCCAGAAAGAAATCGCTTTTAAATCAATCCTGAGTTGATTATTTCACGCTCTTC 2479  
 Db 1181 ---LeuSerAspMetLysLysSerLeuSerAspSerAspAsnValIleSerValIleArg 1199  
 QY 2480 ACTGAGTGTGGATCACAATGAATGCTAATATATCAGAGTACTGAGCAGTACAGTAA 2539  
 Db 1200 SerAspLeuValArgValAsnAspGluLeuAspThrLeuLysLysAspLysAspSerLeu 1219  
 QY 2540 TACACAGGCTATGCTGAGACATATCGTTGTCGCCGAGCCACGAGATGGAGTGAAGG 2599  
 Db 1220 SerThrGlnTyrSerGluValCysGln-----AspArg 1230  
 QY 2600 GAGCAC-----TGCGAGCGGCGC----- 2617  
 Db 1231 AspAspLeuLeuAspSerLeuLysGlyCysGluGluSerPheAsnLysTyrAlaValSer 1250  
 QY 2618 -----TGCACT----- 2623  
 Db 1251 LeuArgGluLeuCysThrLysSerGluIleAspValProValSerGluIleLeuAspAsp 1270  
 QY 2624 -----GAGATCGTTGCCATCGATGCTCTCACTTC 2653  
 Db 1271 AsnPheValPheAsnAlaGlyAsnPheSerGluLeuSerArgLeuThrValLeuSerLeu 1290  
 QY 2654 AGACGCTACCTCGATCAGTTT 2674  
 Db 1291 GluAsnTyrLeuAspAlaPhe 1297

N:Alternate names: hypothetical protein P1760  
C:Species: Saccharomyces cerevisiae  
C>Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 19-Apr-2002  
C:Accession: S65236  
R:Rieger, M.; Mueller-Auer, S.; Schaefer, M.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S65202  
A:Accession: S65236  
A:Molecule type: DNA  
A:Residues: 1-1183 <RIE>  
A:Cross-references: EMBL:Z73573; NID:g1370449; PID:e246934; PID:g1370450; GSPDB:GN00016;  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Gene: SGD:BMS1; MIPS:YPL217C  
A:Cross-references: SGD:S0006138  
A:Map position: 16L  
C:Keywords: transmembrane protein  
F:126-142/Domain: transmembrane #status predicted <TM>  
Alignment Scores:  
Pred. No.: 0.00311 Length: 1183  
Score: 159.50 Matches: 156  
Percent Similarity: 31.89% Conservative: 116  
Best Local Similarity: 18.29% Mismatches: 274  
Query Match: 2.21% Indels: 307  
DB: 2 Gaps: 36  
US-09-302-812-3 (1-4069) x S65236 (1-1183)  
QY 263 AGCAGCAGAGCGCGCTCTCCAGCCCAAGGAC----- 295  
DB 341 ThrArgArgGlySerGluLeuAspLysAspLysLeuIleTyrAlaProMetSerAsp 360  
QY 296 -----GCTCAGTCAGTTCAGGTCGCCACCGTCC 325  
DB 361 ValGlyGlyValLeuMetAspLysAspAlaValTyrIleAspIleGlyLysLysAsnGlu 380  
QY 326 TCGCAGCGCTCGCTCCAGCGGAG-----GCGGACAGCAGCAGGAGCGCCACCTCGCTT 382  
DB 381 GluProSerPheValProGlyGlnGluArgGlyGlyGlyLysLeuMetThrGlyLeu 400  
QY 383 GTTTTCAACAAGAAGACTATTACAGTTCGATGGACACTAAAGGAATC----- 430  
DB 401 GlnSerValGlnSerIleAlaGluLysPheAspGlyValGlyLeuGlnLeuPheSer 420  
QY 431 -----AAGCAGCGGAATCAGAAAGTTTGGATAGTAAGAAACAAC 472  
DB 421 AsnGlyThrGluLeuHisGluValAlaAspHisGlyMetAsp---ValGluSerGly 439  
QY 473 AATACAAAGATAGAA-----TCCATGATGAGTTCTGTACAAAAGAT 514  
DB 440 GluGluSerIleGluAspAspGluGlyLysSerLysGlyArgThrSerLeuArgLysPro 459  
QY 515 ACTTTTACACATAATGTA---GAAAATAGTAATAGTTTCTCAGCTAAGTCTTGAT 571  
DB 460 ArgIleTyrGlyLysProValGlnGluGluAspAlaAspIleAspAsnLeuProSerAsp 479  
QY 572 AGTCACTCACTGAAAAGATACACAGTATTGAAACCAGCATCAGATCGACCAATGTGT 631  
DB 480 GluGlu-----ProTyrThrAsnAsp----- 486  
QY 632 AGTGGCAAAATGAAGGAAACACAGCGAGCTTTTGGAAAGTGAACCTCAACAGTA 691  
DB 487 -----AspAspValGlnAspSerGluProArgMetVal 497  
QY 692 ACCCTGGTACCAGAGAGTTTAGTAATGCTAAACATTGATCGGTGCTCACTCAAAATGATGAT 751  
DB 498 GluIle-----AspPheAsnAsnThrGlyGlnGlyAlaGluLysLeuAlaLeu 514  
QY 752 CACAGTGACACAGATAGTAGAAGATAGAGCAATCAACAGTTTCTCAACAGTGAAG 811  
DB 515 GluThrAspSerGluPheGluGluSerGluAspGluPheSerTrpGluArgThrAla---- 533

QY 812 CTTGCAAAATGCAAGCAGACTACGGAAGATGAACACGCC----- 850  
DB 534 ---AlaAsnLysLeuLysThrGluSerLysLysArgThrTyrAsnIleGlyLysLeu 552  
QY 851 -----AGAGAAGCCAAAGCCACAGAAAGTCCAGCAAGTCTTCCATCTCTGGGAAGAC 904  
DB 553 IleTyrMetAspAsnIleSerProGluGluCysIleArgArgTyrParg---GlyGluAsp 571  
QY 905 TGTGCAAGTTGTCAACAAGATGAGATACAGCTGGTGCCAAAGAGTCCATTGTTCAGATGT 964  
DB 572 AspAspSer-----LysAspGluSerAspIle----- 580  
QY 965 GGCTCTGAGGATGTT-----GGTACTGGGTCAAAAAT 997  
DB 581 ---GluGluAspValAspAspAspPhePheArgLysLysAspGlyThrValThrLysGlu 599  
QY 998 GACAACAAA-----TTGATTAGACAAAGAGTTCCCTAGGAATCTCTCCATTGAG 1051  
DB 600 GlyAsnLysAspHisAlaValAspLeuGluLysPheValProTyrPheAspThrPheGlu 619  
QY 1052 AAGGAAAGTGAACCCGAATCACCGATGGATG----- 1084  
DB 620 LysLeuAlaLysLysTyrLysSerValAspAlaIleLysGluArgPheLeuGlyAlaGly 639  
QY 1085 -----GATAATTCTTAAAAATAGTTTGTCAAGACTCAGAAGCAGATCAGGAGACAAGT 1135  
DB 640 IleLeuGlyAsnAspAsnLysThrLysSerAspSerAsnGluGlyGlyGluLeuTyr 659  
QY 1136 CCAGGTTTGTATGAACAAGATGTTAGTCTCTCCAAACAGCAATAAATCAAGG 1195  
DB 660 GlyAspPheGluAspLeuGluAspGlyAsnProSerGluGlnAlaGluAspAsnSerAsp 679  
QY 1196 TTCCAAGCAAGAGAGCGCTGAC----- 1216  
DB 680 LysGluSerGluAspGluAspGluAsnGluAspThrAsnGlyAspAspAsnSerPhe 699  
QY 1217 -----ATTGAATTTAGGAACCGGTAC 1237  
DB 700 ThrAsnPheAspAlaGluGlyLysLysAspLeuThrMetGluGlnGluArgGluMetAsn 719  
QY 1238 TCTACTAAGCGCGGTGAAGTTAGATTACATTTCCATTGAAGGAGGAGAGATCGCACT 1297  
DB 720 AlaAlaLysLysGluLysLeuArgAlaGlnPheGluIleGluGluGlyGluAsnPheLys 739  
QY 1298 GGAATCAATGATTAAAT-----GCTAACTACCT 1327  
DB 740 GluAspAspGluAsnAsnGluTyrAspThrTyrGluLeuGlnLysAlaLysIleSer 759  
QY 1328 GGAATATTTCTAGCGCTGAATGTAGATGC-----AGAAATTTCT 1366  
DB 760 LysGlnLeuGluIleAsnAsnIleGluTyrGlnGluMetThrProGluGlnArgGlnArg 779  
QY 1367 AAGCAACATGAAAAGAGATCTTAAATCACAGATCAATTCATGATGAGATGCCCC----- 1420  
DB 780 IleGluGlyPheLysAlaGlySerTyrValArgIleValPheGluLysValProMetGlu 799  
QY 1420 ----- 1420  
DB 800 PheValLysAsnPheAsnProLysPheProIleValMetGlyLysLeuLeuProThrGlu 819  
QY 1421 -----AAGCAGAGGACAGAAAGAAAGAACAGTGGGAA----- 1453  
DB 820 IleLysPheGlyIleValLysAlaArgLeuArgArg---HisArgTrpHisLysIle 838  
QY 1453 ----- 1453  
DB 839 LeuLysThrAsnAspProLeuValLeuSerLeuGlyTyrArgPheGlnThrLeuPro 858  
QY 1454 -----ACCAACATCAAAAGAAAGAAAGATCCCTAAATACGTTCCACTCAC 1504  
DB 859 IleTyrThrThrAspSerArgThrArgMetLeuLysTyrThrProGluHis 878  
QY 1505 CTTTCTCCAGATAAGAAGTGGCTTGA-----ACTCCCATTCAGAGAGATGAGAAGA 1555



```
QY 902 GACTGTGCAAGTTGTGTCAGCAAGATGAGATGAGTGGTGGTCCAAAGAGTCCATTGTTCAGAT 961
Db   ::::|||||
QY 227 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 246
Db   ::::|||||
QY 962 GTTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAACAATTTGATTAGACAAGAA 1021
Db   ::::|||||
QY 247 SerGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 262
Db   ::::|||||
QY 1022 AGTTGCCTAGGAAATTTCTCTCCATTGAGAGGAAAGTGAACCCGAATCACCAGTGGAT 1081
Db   ::::|||||
QY 263 SerThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 282
Db   ::::|||||
QY 1082 GTCGATAATCTAAAATAGTTCTCAAGACTCAGAGCAGATGAGGAGACAGTCCAGGT 1141
Db   ::::|||||
QY 283 LeuGluThrLysGluAlaThrAlaAspGluSerLysAlaGluGluThrProIaSerSer 302
Db   ::::|||||
QY 1142 TTGTGATGAACAGAGAGATGGTACTCTCCCAACAGCAAAATAAACCTTCAAGGTTCCAA 1201
Db   ::::|||||
QY 303 AsnGluSerThrProSerAlaSerSerSerSerSerSerSerSerSerSerSerSerSerSer 322
Db   ::::|||||
QY 1202 GCAAGAGACGCTGACATTTGAATTTAGGAACGGTACTCTACTAAGGGCGGTGAGTTAGA 1261
Db   ::::|||||
QY 323 GlyThrAspGluIleLysGluGlyGlnArgLysHisPheSerArgValAspArgSerLys 342
Db   ::::|||||
QY 1262 TTACATTTCCAA 1273
Db   ::::|||||
QY 343 IleAsnPheGlu 346
```

Search completed: May 26, 2004, 17:17:01  
Job time : 212.062 secs







```

QY 2440 GATTGGCAGCGCGCTGTCACGAGATCGTTGCCATTGACGCACTTCACTTCAGACGCTAC 2499
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 -----ThrArgIleValAlaIleAspAlaLeuArg----- 253
QY 2500 CTCGATCAGTTTGTGCTGAGAAAGTGAGACGTGAGCTTAACAAGCTTACTGCGGATTC 2559
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
253 ----- 253
QY 2560 CTCGCTGCTGGAGTTCCTTCTGAAATCTTCTGCAATCTTCTGCACTGGCCAGCAACTGGGCTGT 2619
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 ---HisProGluVal----- 257
QY 2620 GGTGCTTTGGGGTGACGCTAGATTAAAGCCTTAATACAGATCTGTCAGCTGCTGG 2679
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
258 -----SerGlnTyrLysLeuGluSerLeuLeuIleLeuSerVal 273
QY 2680 GCTGAACGTGACGTGGTTTATTCACC-----TTTGGGACTCA 2718
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 SerGlyArgProIleArgLeuTyrMetGlySerValSerLeuGlnGlyIleGlyAspVal 293
QY 2719 GAGTTGATCAGACATCTTACAGATGCACACTTCTTCTTACGAGAGGAGCTGGATGT 2778
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
294 ValLeuMetValGluLeuLeuSerSerLeuPhePheAsnGlyLeuArgPheHisArg 313
QY 2779 GGAAAAGTGCTAC 2790
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 SerAsnLeuTyr 317

RESULT 4
T29757
protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C:Accession: T29757
R:Du, Z.; Le, T.F.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C09D1.
A:Reference number: Z20679
A:Accession: T29757
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6642 <DUZ>
A:Cross-references: EMBL:AF003131; PIDN:AA854132.1; GSPDB:GN00019; CESP:unc-89
A:Experimental source: strain Bristol N2; clone C09D1
C:Genetics:
A:Gene: CESP:unc-89
A:Map position: 1
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Alignment Scores:
Pred. No.: 0.000461 Length: 6642
Score: 176.00 Matches: 187
Percent Similarity: 34.46% Conservative: 120
Best Local Similarity: 20.99% Mismatches: 372
Query Match: 2.60% Indels: 212
DB: 42 Gaps: 42

US-09-302-812-5 (1-3814) x T29757 (1-6642)

QY 88 TCTGCGCCGCTGCTCGGAC---TCCGAGGCTTCCT-----GGCAGCAG 132
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1438 SerSerProThrLysLysGluLysSerProSerProThrLysLysThrGlyAspGlu 1457
QY 133 AGGCGTCTTCGACCCCAAGGACGCTCCGTCAGTTCAGGTCCTCGTCTCTCGCCCA 192
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1458 ValLysGluLysSerProProLysSerProThrLysLysGluLysSerProGluLysPro 1477
QY 193 GCCTGCTCTCGGCGCGGCGGACCGCACAGGC---AACGCCACTCGTTTGT--- 246
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1478 GluAspValLysSerProValLysLysGluLysSerProAspAlaThrAsnIleValGlu 1497

```

```

QY 247 -----TTCAACACAAAGACTATTACTTACTTGGATGATACT 282
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1498 ValSerSerGluThrThrIleGluLysThrGluThrThrMetThrThrGluMetThrHis 1517
QY 283 AAAGGACCAAGACAGCTGAATCAAGAAAGTAAAGAAACAAACAATACAGAATGACTCC 342
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1518 GluSerGluSerArgThrSerValLysLysGluLysThrProGluLysValAspGlu 1537
QY 343 ATGATGATCTTGTGCAGAAAGATACATTTTACCACCATNAG-----GTGCAAAATG 396
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1538 LysProLysSerProThrLysLysAspLysSerProGluLysSerIleThrGluGluIle 1557
QY 397 GAAATGTTCTCAGCTAAATCTTGATAAATCAACCCACA-----GAAAAGAGTTCACAG 450
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1558 -----LysSerProValLysLysGluLysSerProGlu 1568
QY 451 TATTTGAACCAACAGCAGACTGCGAGTGTGCAAGTGCAGATGCAAGGAAAGCATGCA 510
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1569 LysValGluGluLysProAlaSerProThrLysLysGluLysSerProGluLysProAla 1588
QY 511 GAACAGCTTTGGCAAGTGCCTCCCGGGGACTCCGCTA----- 552
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1589 SerProThrLysLysSerGluAsnGluValLysSerProThrLysLysGluLysSerPro 1608
QY 553 -----CCAAAGCAGCTTAGTAAATCTGAACATTGTT 582
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1609 GluLysSerValValGluLeuLysSerProLysGluLysSerProGluLysAlaAsp 1628
QY 583 CAGTCACCCCACT-----GAT 600
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1629 AspLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648
QY 601 GACCACAGTGAACAGATCATCAAGAA-----GACAGAGACAATCAGCAGTTCTTACA 654
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668
QY 655 CTTATATAACTTGCATAACAAAGCAACAGTAGGAGATGGCAGCCAGCAAGCAACTGT 714
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1669 ProThrLysLysGluSerSerProThrLysLysThrAspGluValLysSerProThr 1688
QY 715 AAGTGCAGTGATCTCGCCAGTCTGTGAAGAAGCTGTACAGGTGTCAACAGGAGGAGTG 774
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1689 LysLysGluLysSerProGlnThrValGlu----- 1698
QY 775 GATGCTCTACGAGAGTCTCTTTCAGATGTTGGTCCGAGGACATTTGGAAGTGGACCA 834
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1699 -----GluLysProAlaSerProThrLysLysGlu----- 1708
QY 835 AAAAATGACACAAATTTGACTGGACAAGAA---AGCAGCCTAGGTGATTCGCTCCATTT 891
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1709 LysSerProGluLysSerValValGluValLysSerProLysGluLysSerProGlu 1728
QY 892 GAGAAAGAAAGTGAGCTGAGTCACCAATGGATGTAGACAACTCGAGAAACAGTTGTCAA 951
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1729 LysAlaGluLysProLysSerPro-----ThrLysLysGluLysSerProGlu 1745
QY 952 GATTCAGAGCAGATCAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTTCTCC 1011
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1746 LysSerAlaAlaGluValLysSerProThrLysLysGluLysSerProGluLysSer 1765
QY 1012 -----CAAACAGCAATAAATCTTCAAGTTGCCAAGCAAGAGAGCTGAT 1056
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1766 AlaGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785
QY 1057 GCGATCTTAGGAAACGGTATTTGACTAAGGGAAGTGAAGTAGATTAGATTCATTTCCAATTT 1116
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu----- 1802
QY 1117 GAAGGAAATTAATCTCTGGACCACTGACTTAATGCAAGCCATCTGGAATCTTCT 1176
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1803 -----GluLysProAlaSerProThrLysLysGluLysThrProGluLysSerAlaAla 1820
QY 1177 AGCCTTAATGTAGAGTGTAGAGTTCCAGCAGCATGGAAGGATTTCTAAATTTACA 1236

```

Db 1821 Glu-----GluLeuLysSerProThrLysLysGluLysSerProSerSerProThr 1837  
 QY 1237 GATCATTTTCATGAAATTTCCAAAGTCAGAGGACAGAGAAAGAAACAATGT---GAAGTC 1293  
 Db 1838 -----LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851  
 QY 1294 AGACATCAAGAACAGAAAGAGATCCAAATATACATCCACCTACCTCCCTCCAGAG 1353  
 Db 1852 ProGluLysProLysSerProThrProLysLysSerProGlySerProLysLys 1871  
 QY 1354 AAGAGTGGCTGGAACTCTATTGAGAAATGAGAAATATGCTCGGTGGGATCCAT 1413  
 Db 1872 LysLys-----SerLysSerProGluLysLysPro----- 1883  
 QY 1414 TTGCTTCTTAAAGACCATCTGCAAGTCACACAGTGTCTGCGGTAGACCTTCTGAGA 1473  
 Db 1884 AlaProLysLeuThrArgAspLeuLysGlnThrValAsnLysThrAspLeuAlaHis 1903  
 QY 1474 GCAGGAGGTTCCGAACCTTTTCCACACATTTACAAAGATTTGTGGGATACAAACAT 1533  
 Db 1904 Phe-----GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLysGlu 1922  
 QY 1534 GTGAAATGCTTGTTCGGAACAAACTTGTACCTCTGGAAGATGAGAAATGTGACGGA 1593  
 Db 1923 Ile-----ThrThrAlaGlnGlyValThrValSerLysAspGlnPheGluPhe 1939  
 QY 1594 ACTGCAGGAGTAGGTGGAGCTCATTCAGACTGCACTTTCACAAATTCACAGACCC 1653  
 Db 1940 ArgCysSer-----IleAspThrThrMetPheGlySerGlyThr----- 1952  
 QY 1654 CAGAACTTGAAGTGGATTTCTGAATACATATGTCATATTTCT----- 1698  
 Db 1953 -----ValSerValValAsnSerAsnAlaLysSerValGluThrLysThr 1968  
 QY 1699 -----AAGAAATGGGACTTTACA----- 1716  
 Db 1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988  
 QY 1717 -----GCTTTGGTGTATTTCTGGATAAGTACTTGAAGACGAGGCGCCAAACAT--- 1767  
 Db 1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008  
 QY 1768 -----TTATATCAG-----TCC 1779  
 Db 2009 ProLeuThrLysTrpTyrGlnAsnGlyAsnLeuLeuGluAspGlyLysAsnGlyValThr 2028  
 QY 1780 ATTTTACTGACATGGTGAAATTCGACTCTGTCTGCAAAATATTTCACCCAGCCCAATA 1839  
 Db 2029 IleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn-----AlaGlnAspSer 2046  
 QY 1840 CCATCTCTGAACAGAGATGAATCATCTCTGTCAGATGTCACAGGAAACAGATGCCAGT 1899  
 Db 2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSerAlaGln 2065  
 QY 1900 CTTTATAGCTAATGCTTTCTCTGACACA-----TTTCCCGCA----- 1935  
 Db 2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085  
 QY 1936 -----CGAAATGCCAAGATGAATCGGAGTATTCTAGTTACCCA----- 1974  
 Db 2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105  
 QY 1975 -----GACATTAATCTCAATCGGTGTTTGAAGACGTTTCATCAGGAACACAGAA 2025  
 Db 2106 ValLysTrpThrIleAsn---GluLysIleValGlu-----GluSerArgThrIleThr 2122  
 QY 2026 AAACCTGAACACTCTTCTGCTACTTTTCGAGAGTCACAGAGAAAGAA-----CCTACA 2079  
 Db 2123 ThrIleLysThrGluAspValThrThrLeuLysIleSerAsnAlaLysIleGluLysThr 2142  
 QY 2080 GGATTTGGTGACATTTTACAGACAGAGT-----CTTGAAGATTTT 2118

Db 2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162  
 QY 2119 CCAGATGGGAAGGTGTGAAAAGCCTCTCACACGGCTTACACCTTACGAGGGTACC 2178  
 Db 2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182  
 QY 2179 ATAGAAGGCAACGCGC---CGAGGCATGCTACAGGTGGATTTTCCAAATCGTTTGTGTGGA 2235  
 Db 2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp----- 2195  
 QY 2236 GTGGGTGTGACTGTGTGCGGACTTGTACAGAAGAAATCAGATTTTAATCAATCCTGAA 2295  
 Db 2196 GlyProSerProGlyThr-----GluValSerTrpLeuLeuAsnGlyGln 2210  
 QY 2296 TTGATTTTTCACGGCTGTTCTAGTGTGCTGCATCACAAATCAG---TGCTCTATTATC 2352  
 Db 2211 ProLeuThrLysSerAspThrValGlnValValAsnHisGlyAspGlyThrThrHisVal 2230  
 QY 2353 ACAGGTACTGAACAGTACAGTGTGAATACACAGGC 2385  
 Db 2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241  
 RESULT 5  
 T43523  
 cut17 protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text\_change 02-Jun-2000  
 C:Accession: T43523; T41649; T41700  
 R:Moriwaka, J.; Matsusaka, T.; Yanagida, M.  
 submitted to the EMBL Data Library, August 1999  
 A:Description: Fission yeast cut17 is required for chromosome segregation.  
 A:Reference number: Z22536  
 A:Accession: T43523  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-997 <MOR>  
 A:Cross-references: EMBL:AB031034; FIDN:BA083415.1  
 R:Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, August 1998  
 A:Reference number: Z22007  
 A:Accession: T41649  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-997 <HAR>  
 A:Cross-references: EMBL:AL031323; FIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC962.02c  
 R:Wedler, H.; Duesternoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, October 1999  
 A:Reference number: Z22010  
 A:Accession: T41700  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 932-997 <WEB>  
 A:Cross-references: EMBL:AL121859; FIDN:CA58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10c  
 C:Genetics:  
 A:Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c  
 A:Map position: 3L  
 A:Introns: 43/3  
 Alignment Scores:  
 Pred. No.: 0.000691 Length: 997  
 Score: 171.00 Matches: 188  
 Percent Similarity: 34.44% Conservative: 133  
 Best Local Similarity: 20.17% Mismatches: 344  
 Query Match: 2.52% Indels: 267  
 DB: 2 Gaps: 45  
 US-09-302-812-5 (1-3814) x T43523 (1-997)  
 QY 46 TGGGAGGCC-----TGACGAAA 63  
 Db 173 TrpGluProAspAspAspProTyrThrGluHisLysArgArgAlaAspCysValPhe 192

QY 64 GCGCGCTGGGGCCGCTGGAACTTCTGCGCGCAGCTGCTCGGACTCCCGAGCTTCCCT 123  
Db |||  
193 PheThrTyrLysAspProAsnSerLeuSerProThr- - - - -LysLeuSerPheLeu 209  
QY 124 GCGCGCAGAGCGGTCTTCGACCCCGAGGAGCGCTCCCGTCCAGTTCAGGGTCCCTCG 183  
Db |||  
210 SerThrSerAsn- - - - -IleAspProGluAspLeuThrGluAspAsnSerIleLeuPro 227  
QY 184 TCCTCGCAGCGCTGCTGCTCGGGGGCGGCGGACGCGCACAGAGGCAACGCCACCTCGTTT 243  
Db |||  
228 ValSerPro- - - - - 230  
QY 244 GTTTTCAAAACAAAGACTATTACTCTGGATGGATGATCTAAAGGACCCAGACAGCTGAA 303  
Db |||  
231 - - - - -ThrArgAspSerThrLysSerHisLysThrLeuAsn 242  
QY 304 - - - - -TCAGAAAGTAAGAAACAC- - - - -AATACAGAAATTGACTCCATGATG- - - - - 348  
Db |||  
243 PheSerProSerArgLysAsnAsnLeuAsnAlaArgProLeuThrMetSerLeuTyrThr 262  
QY 349 - - - - -AGTTCTGTGCAGAAAGATAACTTTTACCCACATPAAGGTGGGAAATTTGAAAATGTT 405  
Db |||  
263 AsnThrSerGluGluLysAspSer- - - - -GlnProThrArgAlaProGlnSerProThrLys 281  
QY 406 CCTCAGCTA- - - - -AATCTGTATAATACCCACATACCCACAGAAAGATTACAG 450  
Db |||  
282 ProValLeuLeuThrAlaProArgArgLysAsnLysSerPro- - - - - 295  
QY 451 TATTTCAGCAACACAGCAGACTCGAGTGTGTGCAAG- - - - -TGGCAGAAAT 495  
Db |||  
296 - - - - -LysLysSerLysProAlaValPheLysProValLysProIlePheSerAsp 312  
QY 496 GAAGGAAGCATGCAGACAGCTTTTGGCAAGTGAGCTCCCGCGGGAGCTCGCTACCA 555  
Db |||  
313 GluAspGluAspAspAspLeuThrAlaSerGln- - - - -ProPheSer 327  
QY 556 AGCAGCTTAGTAAT- - - - -GCTAATCTGTGTCAGTACCCACACTGATGACCACT 609  
Db |||  
328 LysGlyIleCysAsnAspSerMetGlnValAlaLysLysAsnPheThrGluGluLeuPro 347  
QY 610 GACACAGATCATCAAGACAGACAGACAATCAGCAGTTCTTACACTATAAACTTGCA 669  
Db |||  
348 LeuLysGluAspGluLysAspAsnGluLeuGluHisLeuValSerProAlaThrSerVal 367  
QY 670 AATACAAAGCCACAGTAGGAGATGGCGAGGCCAGAGCAACTGTAAGTGCAGTGATCT 729  
Db |||  
368 HisThr- - - - - 369  
QY 730 CGCCAGTCTGTGAAGACTGTACAGGTGTCAACAGAGGAGGTGGATGTGCTACCAGAG 789  
Db |||  
370 - - - - -ThrValSerAspIleThrGlyHisGlnSer- - - - - 379  
QY 790 AGTCCCTTGTTCAGATGTGTGTCGCGAGGACATTTGGAAGTGGACCAAAAATGACACAAA 849  
Db |||  
380 - - - - -ValThrAspGluSerAspGlu- - - - - 386  
QY 850 TTCACTGGACAGAAAGACGCTAGGTGATTCGCTCCATTTGAGAAAGAAAGTGCCT 909  
Db |||  
387 - - - - -GlnAsnAsnCysMetSerThrProLysIleGluLeuSerLysIle 403  
QY 910 GATGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 969  
Db |||  
404 GluGluGluIleSerValValSer- - - - -LysSerLysGluIleSerSer 418  
QY 970 GAAACAGTCCAGTCTTTGATGAGCAAGATGATGATGATGATGATGATGATGATGATGAT 1029  
Db |||  
419 SerValSerSerValGlyLysGluGlnAsnHisThrGluLysGlnValAlaIleGluThr 438  
QY 1030 TCAAGTTGCCAAGCAGACAGAGAGCTGATGCGCATCTTAGGAAACGGTATTGACTAAGGGA 1089  
Db |||  
439 ProGluGlnGlnLysValGluLysGluAspGlu- - - - - 449  
QY 1090 AGTGAAGTTAGATTGCATTTCCCAATTTGAAGGAGAA- - - - -AATATGCTGGGACCACT 1143

Db |||  
450 - - - - -HisLeuAsnLeuGlnGlySerPheIleGluSerThrLysGln 464  
QY 1144 GACTTAAATGCCAACCTCTGGAACCTCTTACCTTAATGTAGAGTGTAGAGTTCC 1203  
Db |||  
465 ProfileSerLysProSerThrSerSerProAspMetThrAspAlaIleThrGlyGly 484  
QY 1204 AAGCAGCATGNA- - - - -AGGATCT- - - - -AAA 1230  
Db |||  
485 ArgValSerSerSerPheArgAspLysIleLeuGlnThrAsnPheSerProArgSer 504  
QY 1231 ATTACAGATCATTTTCATGAGATTTCCAAAGTCAGAGGACAGAGAAACAAATGTGAA 1290  
Db |||  
505 ThrIleAspSerPheSerAsnIleSerLysLys- - - - -ArgAsnSerGluGluAlaAsn 522  
QY 1291 GTCAGACATCAAGAACAGAAAGGAGATTCCAAATATACATCCACCTACCTCCCTCCA 1350  
Db |||  
523 AspGluAsnAspGluThrAsnLeuLysIlePro- - - - -IlePro 535  
QY 1351 GAGAGAAATGGCTGGAACTCTTATGAGGAAATGAGAAATGCTCGGTGGGATC 1410  
Db |||  
536 GluLysLysArgLysPheGlnGluValLeuGlnSerLysAsnIle- - - - - 550  
QY 1411 CATTTGCTCTTAAAGACCATCTGCAGTCACAGTCACAGTGTGTTGGGTAGACCTTCTG 1470  
Db |||  
551 - - - - -LeuValSerSerThrGluAspSerHisGluProValLysValThrGluAspSerGln 569  
QY 1471 AGAGCAGGAGAGTTCGGAACCTTTTCCACACATTCACAGATTTGGGATACAAA 1530  
Db |||  
570 ThrAlaIleHisValSerLys- - - - -PheGluAspLeu- - - - -GluAsnLys 583  
QY 1531 CATGTGAAAATCCCTTGTTCGGAACAAACTGTGACCTGTG- - - - -GAAGATCAGAAATGCT 1587  
Db |||  
584 SerMetGlu- - - - -SerGluGlnSerLeuGlnLeuLeuSerGluSerGluAsnAsp 600  
QY 1588 GAGCGAACTGCAGGAGTAGTGGAGCTCATTCAGACTGACCTCTCTCAACAAATTCACA 1647  
Db |||  
601 AspLys- - - - -ProLeuIleAspLeuIleProLeuLeuAlaIleLys 614  
QY 1648 CGACCCCGAAGTGAAGATGCGATTCTGAAATACATATGTCGCTATCT- - - - - 1698  
Db |||  
615 ArgLysAspAsnLeuValSerGlyValLeuGluLysGlyLysSerThrSerLys 634  
QY 1699 AAGAAATGGAGCTTACAGCTTTGTTGATTTCTGGGATAAGTACTTGAAGAGCAGAG 1758  
Db |||  
635 ThrLysPheAsp- - - - -ThrSerIleValAspPheIleGluLysProLysThrGlu- - - - - 651  
QY 1759 GCCCAACATTTATATCAGTCCATTTTACCTGATGATGCTGAAATTCACCTCTCTGCGCA 1818  
Db |||  
652 - - - - -IleSerGluValLeuProGluGluLysArgLysAlaIleCys- - - - - 665  
QY 1819 AATATTGACCCAGCAATACCATCTCTGAAACAGAGATGATCATCTCTGTCAGATG 1878  
Db |||  
666 - - - - -AspGluSerGlnThrValArgValSerIleAspArgGlyValThrLys 681  
QY 1879 TCACAGGAACAGATCCGAGCTCTTTTAGCTAATGCTTCTCTGCACTTCCCGACCG 1938  
Db |||  
682 ThrArgAsp- - - - -ValSerSerProValSerAspGlu- - - - -LysSerGlu 695  
QY 1939 AATGCCAAGATGAAATCGGAGTATTCTAGTTACCCAGACATTAATCTCAATCGGTTGTTT 1998  
Db |||  
696 AsnValAsnHisGluGluAlaAsnSerGlyHisThrValMetAsnValHisSerSerLeu 715  
QY 1999 GAAGGAGCTTCACAGGAAACCCAGAAACCTGAAACACTCTCTCTGCTACTTTCGAAGA 2058  
Db |||  
716 AspProGlnProIleValGlnProAsnGluLeuLysSer- - - - -GlySerThrLeuLysAsp 734  
QY 2059 GTCACAGAGAAAAA- - - - -CCTACAGGATTTGGTACATTTTCAACAGACAGACTCTTGA 2112  
Db |||  
735 LeuProAspArgAsnValGlyAsnSerGluLysValThrPheGlnGluAspAspIleAsn 754  
QY 2113 GATTTTCCAGATGGGAAAGGTGTGAAAGCCCTCTGACAGGCTTACAGCTCATTAGCAG 2172  
Db |||





```
Db 234 LysAlaAlaAlaThrProLysLysThrValProLysLysGlnValValAlaLysAlaPro 253
Qy 427 -----TCACCCACAGAAAGATTTCACAGTATTGTAACCAACACAGACT 471
Db 254 ValLysAlaAlaThrThrProThrArgLysSerSer-----SerSerGluAspSer 270
Qy 472 GCGAGTGTGTCAGTGCAGAAATGAGGGAAGCATGCAGAAACAGCTTTTGGCAAGTGAG 531
Db 271 SerSer-----AspGluGluGluGluGlnLysLysProMetLysAsnLys 285
Qy 532 -----CTCCCGCGGGGACTCCGCTACCAAGCAGCTTAGTAAT 570
Db 286 ProGlyProTyrSerTyrAlaProProSerAlaProProProLysLys----- 302
Qy 571 GCTAACATTGCTCAGTCACCC----- 591
Db 303 ---SerLeuGlyThrGlnProProLysLysAlaValGluLysGlnGlnProValGluSer 321
Qy 592 -----CACACTGATGACACAGTGCACAGATCATGAAGAGACAGACAAATCAGCAG 645
Db 322 SerGluAspSerSerAspGluSerAspSerSerSerGluGluGluLysLys----- 338
Qy 646 TTTCTTACACTTAAACTTGCATAATACAAAGCCACAGTAGGATGGCCAGCCAGA 705
Db 339 -----ProProThrLysAlaValSerLysAlaThrLysProProProAlaLys 356
Qy 706 AGCAACTGTAAGTCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGGCTGTCAACAG 765
Db 357 Lys-----AlaAlaGluSerSerSerSerSerSerSerSerSerSer 370
Qy 766 GAGAGGTGGATGCTACACAGAGTCTTTGTGCAGATGTTGGTCCGAGCAGATTGGA 825
Db 371 SerGluAspGluAlaProSerLysPro-----AlaGly 382
Qy 826 ACTGGACCAAAATGACAAATTTGACTGGACAAAGACGCTAGGTGATTCGCT 885
Db 383 ThrThrLysAsnSerSerAsnLys-----ProAlaValThrThrLysSerPro 398
Qy 886 CCATTGTGAGAAAGAGTGAGCTGAGTCACCAATG----- 921
Db 399 AlaValLysProAlaAlaProLysGlnProValGlyGlyGlnLysLeuLeuThr 418
Qy 922 -----GATGTAGACAACTCGAGAAACAGATTCTCAAGATTTCAGAGCAGATGAAGA 972
Db 419 ArgLysAlaAspSerSerSerSerGluGluGluSerSerSerSerSerGluGluLysThr 438
Qy 973 ACAAGTCCAGCTTTGATGACGCAAGATGTCCTTCCCAACAGCAATAAATTTCA 1032
Db 439 LysLysMetValAlaThrLysProLysAlaThrAlaLysAlaAlaLeuSerLeuPro 458
Qy 1033 AGTTGCCAAGCAGAGAGCTGATCGCATCTTAGGAAACGGTATTTGACTAAGGGAAGT 1092
Db 459 AlaLysGlnAlaProGlnGlySerArgAspSerSerSerSerSerSerSerSer 478
Qy 1093 GAAGTTAGATTGCATTCCAAATTTGAGAGAGAAATATGCTGGGACCACTGACTTAAT 1152
Db 479 Glu-----GluGluGluGluLysThrSerLysSerAlaValLys 491
Qy 1153 GCCAAGCCA-----TCTGAAACTCTTCTAGCTTAAATGATAGTGTAGAAGT 1200
Db 492 LysLysProGlnLysValAlaGlyGlyAlaProSerLysProAlaSerAlaLysLys 511
Qy 1201 TCCAAGCAGCATGGAAGGATCTAAATATACAGATCAATTTTCATGAGATTTTCCAAAG 1260
Db 512 GlyLysAlaGluSerSerAsnSerSerSerSerSerSerSerSerSerSer 526
Qy 1261 TCAGAGGACAGAGAAAGAAACAATGTGAGTCAGACATCAAGAACAGAAAGGAGATT 1320
Db 527 GluGluGlu-GluGluLysLeuLysGlyLysGlySerProArgProGlnAlaProLysAl 546
Qy 1321 CCAAATACATCCACCTTACCTCCCTCCAGAGAGAGTGGCTGGGACTCTATTGAG 1380
```

```
Db 546 aAsnGlyThrSerAlaLeuThrAla-GlnAsnGlyLysAlaAlaLysAsnSerGluGluG 566
Qy 1381 GAAATGAG-----AAAAATGCTCGGTGGGTGGATTCATTGCTTCTTAAAGACCATCT 1434
Db 566 luGluGluGluLysLysAlaAlaVal-----ValValS 578
Qy 1435 GCAAGTCACACAGTACTGTTCGGGTAGACCTTCTGAGCAGAGAGGTTCCGAACCC 1493
Db 578 erLysSerGlySerLeuLysLysArgLysGlnAsnGluAlaAlaLysGluAlaGluThr 597
RESULT 8
I39410
AF-4 protein, splice form 3 - human
N:Alternate names: serine/proline-rich PEL protein
C:Species: Homo sapiens (man)
C:Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 29-Aug-1997
C:Accession: I39410
R:Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaani, O.; Kamada, N.; Lane;
Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993
A:Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leu
A:Reference number: AA7440; MUID:93281633; PMID:8506309
A:Accession: I39410
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1210 <NAK>
A:Cross-references: GB:LI3773; NID:9306446; PID:9306447
C:Comment: This protein is one of several to form chimeric fusion proteins with the trit
C:Genetics:
A:Gene: GDB:MLIT2; AF-4; FEL
A:Cross-references: GDB:I36792; OMIM:159557
A:Map position: 4q21-4q21
C:Keywords: alternative splicing
Alignment Scores:
Pred. No.: 0.00162 Length: 1210
Score: 166.00 Matches: 216
Percent Similarity: 33.43% Conservative: 140
Best Local Similarity: 20.28% Mismatches: 402
Query Match: 2.45% Indels: 307
DB: 2 Gaps: 48
US-09-302-812-5 (1-3814) x I39410 (1-1210)
Qy 15 TCGCGG-----TCCGACGATGAGTCGGGCGCGGCTG-----GGA 50
Db 151 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 170
Qy 51 GCCCTGCACGAAAGCGCGTGGGCGCGCTGGAACTTCTGCGCGACTCTCGGACTC 110
Db 171 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 190
Qy 111 CCGAGCTTCCTCGGCGAGCAGAGCGGTGTTCTGACCCCAAGCAGCGCTCCGCTCCAGT 170
Db 190 xAspSerAlaProGluArgGluLeuSerProLeuIleSerLeuProSerPro----- 207
Qy 171 CAGGTCCCTCCGCTCGCGCGCTCGCTCGGCGCGGCGGCGGAGCAGCAGAGCAA 230
Db 208 ----ValProProLeuSerProIleHisSerAsnGlnHisLeuProArgThrGlnG 226
Qy 231 CGCCACCTCGTTGTTTTCAAAACAAAGACTATTACTACTTGGATGGATATAAGGACC 290
Db 226 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 246
Qy 291 CAAG-----ACAGCTGAATCAGAAAGTAAAGAAACAAACATACAGAAATTGCTCAT 344
Db 246 oLysAspLeuAlaValLysValHisAspLysGlu-----ThrProGlnAspSerLe 263
Qy 345 GATGAGTCTGTGGCAGAAAGATAAC-----TTTACCCACATAAGGT 386
Db 263 uValAlaProAlaGlnProProSerGlnThrPheProProSerLeuProSerLysSe 283
Qy 387 GGAATAATTTGAAATGTTCTCCTCAGCTAAATCTT-----GATAATC 428
```

Db 283 rValalaMetGlnGlnLysProThrAlaTyxValArgProMetAspGlyGlnAspGlnAl 303  
Qy 429 ACCCAGAGAAAG-----AGTTCACAGTATTGGAACCAACAGCAGAC 470  
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 321  
303 aProSerGluSerProGluLeuLysProGluAspTy-----ArgGlnGlnTh 321  
Qy 471 T-----CCGAGTGTGTCAAGTGCACAGATGA 497  
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 341  
321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341  
Qy 498 AGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCTCCCGCGGGAGCTCCGCTACCAAA 557  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 361  
341 oSerGlnSerValGluGlnThrTyxSerAsnGluValHisCysValGluGluLeuLys 361  
Qy 558 GCAGCTTAGTAATGCTAACATGCTGCTGAGTCCACCACTGATGACCACTGCAGCAGACA 617  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 379  
361 sGluMetThrHisSer-----TrrProProProLeuThrAlaLeHisThrProSerTh 379  
Qy 618 TCATCAAGAAGACAGACAGCAATCAGCAGTTCTTACACCTATATAAACTTGCAAAATACAAA 677  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 395  
379 rAlaGluProSerLys-----PheProPheProThrLysAspSerGlnHisVa 395  
Qy 678 GCCAACAGTAGGAGATGGCGAGCCAGAGCAACACTGTAGTGCAGTGGATCTGCCAGTC 737  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 408  
395 lSerSerValThrGlnAsnGlnLysGlnTyxAsp-----ThrSe 408  
Qy 738 TGTGAAAGACTGTACAGCTGTCAACAGGAGGAGGTGGATGCTTACACAGAGAT---CC 794  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 428  
408 rSerLysThrHisAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 428  
Qy 795 TTTGTGCAGATGTTGTGTCGCGAGCAATGGAACTGGACCAAAATATGACAAATATTCAC 854  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 440  
428 nLeuSerAsp-----SerGluAsp 440  
Qy 855 TGGCAAGAAAGACAGCTAGGTGATGCGCTCCATTGAGAAAGAGTGCAGCTGAGTC 914  
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 459  
440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu-- 459  
Qy 915 ACCAATGATGTAGCACTCAGAAACAGTGTCAAGATTTCAGAACGATGAGAAAC 974  
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 479  
460 -ProValAlaSerAlaHisSerSerAlaGluSerGluSerThrSerAspSerAspSe 479  
Qy 975 AAGTCCAGCTTTGTATGACAGAGATGATCGTCTCCCAACAGCAAAATAAACTTTCAAG 1034  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 499  
479 rSerSerAspSerGluSerGluSerSerSerSerAspSerGluGluAsnGluProLeuGl 499  
Qy 1035 TTGCCAAGCAAGAGCTGATGGCGATCTTAGGAACCG-----TATT 1079  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 519  
499 uThrProAlaProGluProGluProProThrThrAsnLysTrpGlnLeuAspAsnTriple 519  
Qy 1080 GACTAAGGGAAGTGA-----GTTAG 1100  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 539  
519 uThrLysValSerGlnProAlaAlaProProGluLysProArgSerThrGluProProAr 539  
Qy 1101 AUTGCATTTCCCAATTTGAAGAGAAATTAATGCTGGGACCACTGACTTAATGCCAAG-- 1158  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 559  
539 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 559  
Qy 1159 -----COATCTGGAATCTTCTAGCTT----- 1182  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 579  
559 xLysAspProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 579  
Qy 1183 -----AATGTAGACTGTAGAGTTCC----- 1203  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 599  
579 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 599  
Qy 1204 -----AAGCAG----- 1209  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 619  
599 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 619  
Qy 1210 -----CATGAAAAGGGATTTCTAAATATACAGATCATTTTCATGAGATTTCCAGTTC 1262  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 639  
619 rSerLeuGlnGlyGluArgGluProGlyLeuLeuProTyxGlySerArgAspGlnThrSe 639

Qy 1263 AGAGCAGACAGAGAAAGAAACAATGTGAAGTCAGACATCAAAAGAACAGAAAGAGATTC 1322  
Db :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 659  
639 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaSerAsnGluPr 659  
Qy 1323 AAAATACATCCCACTAACCCTCCCTCCAGAGAAAGAGTGGCTGGGAATCTCTATTGAGA 1382  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 668  
659 oLys-----ProAlaValProProSerSerGlu----- 668  
Qy 1383 AATGAGAAAATCGCTCGGTGGGATCCATTGCTTCTTCTTAAGACCATCTGCAAGTCA 1442  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 680  
669 ----LysLysLysHisLysSerSerLeuProAlaProSer----- 680  
Qy 1443 CACAGTGACTGTCGGGTAGACCTTCTGAGAGCAGGAGAGGTTCGAAACCTTTTCCAAC 1502  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 691  
691 -----LysAlaLeuSerGlyProGluProAlaLysAs 691  
Qy 1503 ACATTACAAAGATTTGTGGGATAACAACATGTGMAATGCTTGTTCGGACAAACTT 1562  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 710  
691 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 710  
Qy 1563 GTACCTGTGCAAGATGAGATGTGACCGAACTGCGAGGAGTAGGTGGAGCTCATTTCA 1622  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 730  
711 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValVa 730  
Qy 1623 GACTGCATCTTCTCAACAAATTCACACGACCCCAACTTGAAGGATCGATTCGAAATA 1682  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 746  
730 lGlnGluAspSerArgLysAspArgLeuProLeuArgAspThr----- 746  
Qy 1683 CAATGTGCATATTCTTAAGAAATGGGACTTTACAGCTTTGTTGTTCTGGGATAAGT 1742  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 748  
747 -----LysLe 748  
Qy 1743 ACTTGAAGAGCAGAGGCCCAACTTATATCAGTCCATTTTTACCTGACATGTTGAAAT 1802  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 765  
748 uLeuSerProLeuArgAspThrProProGlnSerLeu-----MetValLysAl 765  
Qy 1803 TGCACCTGTCTGCCAAATATTGTCACCCAGCCCA----- 1836  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 785  
765 eThrLeuAspLeuLeuSerArgLysProGlnProProGlyLysGlySerArgGlnArgLy 785  
Qy 1836 ----- 1836  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 805  
785 eAlaGluAspLysGlnProProAlaGlyLysHisSerSerGluLysArgSerSerAs 805  
Qy 1837 -----ATACCACCTCTGAAACAGAGATG-----AATCA 1865  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 825  
805 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 825  
Qy 1866 TTCTGTCTACGATG-----TCACAGAAACAGATCGCCAGTCTTTTACTATA 1910  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 845  
825 sLysLysArgLeuLysGluLysSerGlnSerSerSerSerSerSerSerSerHisLy 845  
Qy 1911 TCCTTTCTTCTGCACTTTCCCGAGCGAATCCCAAGATGAATCGAGTATTCTAGTTA 1970  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 865  
845 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 865  
Qy 1971 CCCAGCATTAATCTCAATCGGTGTTTGAAGGACGTTTCATCAAGGAAACCCAGAAAAA-- 2028  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 880  
865 uProProProProValSer-----SerSerSerGlnLysProAlaLysPr 880  
Qy 2029 -----CTGAAAACACTTCTCTGCTATTTCGGAAGATGCACAGAG-----AAAAAAC 2075  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 896  
880 oAlaLeuLysArg-----SerArgArgGluAlaAspThrCysGlyGlnAspPr 896  
Qy 2076 TACAGATTTGTGACATTTTACAGACAGAGTCTTGAAGT-----TTTCCAGATGGGA 2129  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 916  
896 oProLysSerAlaSerSerThrLysSerAsnHisLysAspSerSerLysProLysGlnAr 916  
Qy 2130 AAGGTGTGAAAACCTCTCTGACACGCTTACACGCTCATCTTACGAGGCTACCAT-- 2181  
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 936  
916 gArgValGluGlyLysGlySerArgSerSerSerGluHisLysGlySerSerGlyAspTh 936



```

QY 2182 -----GAAGGCAACGCCGA-----GGCATGTACA 2207
Db 936 rAlaAsnProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGly 956
QY 2208 GGTGATTTTCAAAATCGTTTTTTGGAGGTGGTGTGACTGGTGGGACTT---GTACA 2264
Db 956 nValLysPheAspLysGln-----GlnAlaAspLeuHisMetAr 969
QY 2265 AGAAGAAATCAGATTTTAAATCAATCCGAATGATTGTTTCACGGCTG----- 2313
Db 969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaPh 989
QY 2314 -----TTCACTGAGGTGGATCACAATGAGTGTCTTATTATCAGAGTACTGAACA 2366
Db 989 eLysTyrLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGluSe 1009
QY 2367 GTACAGTGAA-----TACACAGGCTATCTGTAACCT-----TA 2399
Db 1009 rGlnSerSerLysSerAlaTyrSerValTyrSerGluThrValaAspLeuLysPheIl 1029
QY 2400 TCGTTGGCCCGAAGCCATGAGATGGAGTGGAAGACGATGGCAG----- 2448
Db 1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaVa 1049
QY 2449 -----CGGCGCTGCACGGAGATCGTTGCCATTCAGCAGCTTCACCTTCAGACGCTACT 2501
Db 1049 lLeuCysMetArgCysGlnSerIle-----LeuAsnMetAlaMetPheArgCysLysLy 1067
QY 2502 CGATCAGTTTGTGCTGAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2561
Db 1067 sAsp-----IleAlaIleLysTyrSerArgThrLeuAsnLysHisPhe----- 1081
QY 2562 CCGTCTCGAGTTCCTTCTGAAATCTTCTGCAGTGGCCAGGCAAACTGGGGCTGTGG 2621
Db 1082 -----GluSerSerLysValAlaGlnAlaProSerProCysIl 1095
QY 2622 TGCCTTTGGGGT 2634
Db 1095 eAlaSerThrGly 1099

RESULT 9
A17623
probable secreted protein PF0115w - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: A17623
R:Gardner, M.J.; Tettelin, H.; Carnucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A17600; MUID:99021743; PMID:9804551
A:Accession: A17623
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1192 <GAR>
A:Cross-references: GB:AE001373; GB:AE001362; NID:g3845097; PIDN:AACT71813.1; PID:g384509
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0115w

Alignment Scores:
Pred. No.: 0.00189 Length: 1192
Score: 165.00 Matches: 85
Percent Similarity: 36.96% Conservatives: 61
Best Local Similarity: 21.52% Mismatches: 143
Query Match: 2.44% Indels: 106
DB: 2 Gaps: 18

US-09-302-812-5 (1-3814) x A17623 (1-1192)
QY 235 ACTCGTTTGTTCNAACAAAGACTATTACTACTTGATGGATGAT---ACTAAAGGACCC 291
Db 235 SerAspPheMetAsnLysMetLysTyrGlnAsnThrGlnIleAspAsnAsnLysGlyLys 254

```

```

QY 292 AAGACAGCTGAATCAGAAAGTAAAGAAAACAAACAATACAAAGATTGACTCCCATGATGAGT 351
Db 255 LysThrThrAsnThrMetGluLysAsnLysLysAsnLysLysAsnLysLysLysLysLysLysLys 274
QY 352 TCTGTGCAGAAAGATAACTTTT---TACCACATAGGTGGAAATAATTTGGAATAAT---GTT 405
Db 275 ArgLysThrLysGlnAsnTyrLysTyrLysLysGlnAsnGlnAsnLysLysLysLysLysLysLys 294
QY 406 CCTCAGCTAAAT-----CTTGATAAATCACCACACAAA--- 438
Db 295 ProGlnSerLysTyrLysGlnGluArgIleGluLeuLeuAspAspAsnGlyLysGluLeu 314
QY 439 AAGAGTTTCACAGTATTTTGAACCAACACAGAG-----ACTGCGAGT 477
Db 315 LysSerHisLysAsnLysLysGluLysGlyLysGlyLysGlyLysGlyLysGlyLysGlyLysGly 334
QY 478 GTGTGC-----AAGTGGCAGATGAAGGAAGCATGAGCAACAGCTTTTGT 522
Db 335 IleAlaAspLysLysLysLysLysGluGluArgGluThrLysAspGluLysGluLysGluLysAsn 354
QY 523 GCAAGTGAAGCTCCCGCGGAGCTCCCGTACCAGCAGCTTAGTATGCTTAAGCTTAAGCTTGT 582
Db 355 IleGlnGln-----LeuValLysAspValGlnLeuLysLysValGly 368
QY 583 CAGTCACCCAC-----ACTGATGACCAACAGTGAACA 615
Db 369 GluGluThrLysAspAspGluLysGluAspLysGluGlyLysGlyLysGlyLysGlyLysGlyLys 388
QY 616 GATCATGAAGGAAGACAGAGACAATCAGCAGTTTCTTACCTATAAACTTGCAATAC 675
Db 389 AspAspGluGluAspThrAspAspGluGluAspThrAspAspGluGluAspThrSerAsp 408
QY 676 AAGCCACAGTGAAGATGGCGAGCCAGCAGCACTGTAAGTGCAGTGGATCTCGCCAG 735
Db 409 GluGluThrThrGlyAspGlnGlu----- 416
QY 736 TCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGTGGATGTGCTACCAGAGAGTCTCT 795
Db 417 -----AsnLysGluGluThrGluVal----- 423
QY 796 TTGTCAGATGTTGGTGGCGAGGACATTGGAACTGGCAACCAAAATGACAACAATGACT 855
Db 424 -----GGACAAAGAACGAGCTAGGTGATTCGCTCCATTTGAGAAAGAAAGTGGAGCT 909
QY 856 -----GGACAAAGAACGAGCTAGGTGATTCGCTCCATTTGAGAAAGAAAGTGGAGCT 909
Db 429 GluLysAlaGluGluGluLeuGluGluAspLysGluGluSerGluLysAspLysGlu--- 447
QY 910 GAGTCACCAATGGATGTAGACAACCTCGAGAAACAGTTGTCAGATTTCAGAACGAGATGAA 969
Db 448 GluSerGluLysAspLysGluGluSerGluLysAspLysGluLysGluSerGluLysAspLys 467
QY 970 GAAACAGTCCAGTCTTTGATGACAGATGATCGTTCTCCCAACAGCAAAATAAACTT 1029
Db 468 GluLysThr-----GluGluAspGluGluLysThrGluAsp---GluLysGly 482
QY 1030 TCAAGTTGCCAAGCAAGAGAGAGCTGATGGCGATCTTAGGAAACCGGTATTTGCTAAGGGA 1089
Db 483 ThrGluValTyrLysLysGluThrAspValAspGluLysLysGlu----- 497
QY 1090 AGTGAAGTTAGATTGCATTTCCAAATTTGAAGAGAGAAATAATGCTGGGACCAAGTACTTA 1149
Db 498 -----LysGlyGluTyrGlyGluGlyThrAspAspGlu 508
QY 1150 AATGCCAAGCCATCTGGAACTCTTCTAGCCCTTAATGTAGAGTGTAGAGTTTCCAAGCAG 1209
Db 509 GluAspLysGluLysGluGluAspGlu-----GluThrLysValGluGluLys 525
QY 1210 CATGGAAGAAAGGATTTCTAAATAATTACAGATCATTTTCATGAGAAATTTCCAACTCAGAG 1269
Db 526 LysThrGluLysAspGluGluGlyThrAspTyrGluGluAspThrAspAspSerAsp--- 544

```



[illegible]

```

Db      2001  -----LysGluGlnGluArgValLysArgMetTyrGluGluAsnGluAr 2015
QY      2058  AGTCACAGAGAAAAAACCCTACAGGATTGGTGACATTTACACACAGAGCTCTGAAGATTT 2117
Db      2015  glysValGlu-----GluAspAr 2021
QY      2118  TCCAGAATGGAAAGGTGTGAAAGCCTCTGACAGCTTTACAGCTACAGGGGTAC 2177
Db      2021  gArgGluLysGlnArgLysGluGluGluArgGlnArgLeu-AlaAlaAlaThraAlaAla 2041
QY      2178  CATAGAGCAACAGCCGAGGAGCATGC----- 2203
Db      2041  laThrMetAlaThrGlnLysAlaAlaGluAlaLeuLysGlnLysGlnValProArgH 2061
QY      2204  -----TACAGTGGATTTTGCAAATCGTTTCTGGAGGTGGTGCACTGGTGGCGGACT 2258
Db      2061  isGlyPheGlnHisValLeuSerMetMetThrProGlu-----AlaArgSerL 2077
QY      2259  TGTACAAGAGAAATCAGATTTTAAATCAATCCTGAAT 2296
Db      2077  euTyrGluGlnPheProGlyLeuSerSerTyrIleAsn 2089

RESULT 12
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88320
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; PMID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2738 <STO>
A:Cross-references: GB:chr_II; PIDN:CAA93781.1; PID:G3881547; GSPDB:GN00020; CBSP:F07A11
C:Genetics:
A:Gene: F07A11.6
A:Map position: 2

Alignment Scores:
Pred. No.:
Score: 0.00312 Length: 2738
Matches: 163.00
Percent Similarity: 34.63% Conservative: 129
Best Local Similarity: 19.39% Mismatches: 301
Query Match: 2.41% Indels: 253
DB: 2 Gaps: 36

US-09-302-812-5 (1-3814) x E88320 (1-2738)
QY      52  CCTGCACGAACGGCGCTGGGGCGCGTGGNACTTCTGCGCGACTCGCTCGACTCC 111
Db      1416  ProArgGlnArgAsnArgThrThrSerThrSerThrAlaThrThrSerSerLysHis 1435
QY      112  CGAGCTTCCCTCGGAGGAGGCGGTCTCGACCCCAAGGACGCTCCCGTCCAGTTTC 171
Db      1436  GluAlaLeuSerLleProGluLysProLeu-SerProProValThrAlaLysSerSerVa 1455
QY      172  AGGCTCCCTC----- 1738
Db      1455  lSerSerLleAspProSerLleArgAspGluPheSerMetMetAsnSerAlaAlaAspSe 1475
QY      193  GCCTGCTCTCGGGGGGGGGGACCGCA-----GA 225
Db      1475  rProMetSerThrThrGlyArgProMetValLeuThrLysAlaAlaMetLysAlaPheAs 1495
QY      226  GGCAACGCCACCTCGTTGTTTTCACAAACAAAGACATATTA-----CTACTCTGGATG 276
Db      1495  nSerThrProProLysLysValSerTyrSerLleLleAspCysTyrMetLeuGlyMe 1515

```

```

QY      277  -----GATACTAAGAGCCCAAGACACACTGAATCAGAAAGTAAAGAAACAA 323
Db      1515  tMetAlaLysAsnSerSerGlyGlnHisAspSerSerGlySerSerSerAspSe 1535
QY      324  CAATACAAAGATTTGACTCCATGATGATTCGTGCAGAAAGATAACTTTTACCCACATAA 383
Db      1535  rSerSerSerAspGlySerThrSerSerAspAspSerSerAspGluValProLysG 1555
QY      384  GGTGAAAAAATTTGAAAAATTTCTCAGCTAAATCTTGAATAATCACCACAGAAAG-- 441
Db      1555  nThrGluProValThrSerLleProValAlaSerAspAsnGlySerProGluAsnVa 1575
QY      442  -----AGTTCACAGTATTGAAACCAACAGACAGACTGCGAGTGT 479
Db      1575  lValValGluThrProSerLleValSerGlnThrProArgGluProGluProPheThr 1595
QY      480  GTGCAAGTGGCAGAGTAAGGAAGCATGACAGACAGCTTTTGGCAAGTAGAGCTCCCG 539
Db      1595  eSerGluGlnSerSerGlu-----SerGluProGluAlaValProGluCysProGl 1612
QY      540  GGGGACTCCGCTACCAAGCAGCTTTAGTAATGCTAACTTGT-----CAGTCACCCCA 593
Db      1612  uAlaSerValGluProGlnMetGluThrSerGlnAsnValGluProValSerGluGlu 1632
QY      594  CACTGATGACCACTGACACAGATCATGAAAGACAGACAGACATCAGCAGTTCTTAC 653
Db      1632  sGluAspSerHisGluHisGlyAspSerGluValAlaValGluSerGlnGln-----G 1650
QY      654  ACCTATAAACTTGCAAATATACAAAGCCACAGTAGAGATGGCAGGCCAGAGCAACTG 713
Db      1650  nProLeuGlu----- 1653
QY      714  TAAGTGCAGTGGATCTCCGACTCTGTGAAAGACTGTACAGGCTGTCAACAGAGAGGT 773
Db      1654  -----HisGlnGluGlu 1658
QY      774  GGATGTGTACAGAGAGTCTTTTGTGATGTTGTCGCGAGGACATTTGGAACGTGACC 833
Db      1658  sGluGluLeu-----GluAsnLysLleLeuValAlaAlaGlu----- 1671
QY      834  AAAAATGACAAACAAATTTGACTGGCAAGAAAGACGCTAGGTGATTCG---CCTCCATT 890
Db      1672  -HisGlnGluGlnValGlnGlyAspGluAspSerValGluSerSerLleProAlaPr 1691
QY      891  TCAGAAAGAAAGTGAAGCTGAGTACCACCAATGATGTAGACAACTCGAGAAACAGTTG 950
Db      1691  oSerAspGluProAspProValThrGlnAlaGlnGluLysSerAlaHisThrLeuIle 1711
QY      951  AGATTTCAGAGCAGATGAAAGAAACAGTCCAGTCTTTGATGAGCAGAT---GATCGTTC 1007
Db      1711  rAspGlnGluThrAspGlnAlaValGlnSerLlePheAspGluGluGluAlaAspGlu 1731
QY      1008  CTCCTCAACAGCAAAATAAACTTTCAAGTTGCCAAGCAAGAGAGCTGATGGCATCTTAG 1067
Db      1731  eProGlnTyrProAspPheGlyLleSerThrAsnGluLysGluValSerGly-----Ly 1749
QY      1068  GAAACGGTATTGACTTAAGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1127
Db      1749  sAspProHisAsnLleLysProThrGlu-----ProLe 1760
QY      1128  TAATGCTGGACCACTGACTTAATGTCGAAGCCATCTGGAACCTCTTCTAGCTTAATGT 1187
Db      1760  uAsnAsnGlyHisThrAspLeuLysPheSerProSerSerSerAlaHisAla----- 1777
QY      1188  AGAGTGTAGAGTCTCAAGCAGCAGTGAAGAAAGAGGATTTCTAAATTTACAGATCATTT 1247
Db      1778  -----SerGluLysGlnSer----- 1782
QY      1248  GAGAATTTCCAGTGCAGGACAGAGAAAGAAAGAAACATGTGAAGTCAGACATCAAGAAC 1307
Db      1783  -----ThrLysSerGluAspAspMetGluGluAspSerGluLeuValMetGlu 1800

```

QY 1308 AGAAGGAAGATTCACAAATACATCCACCT-----AACCTCCTCCAGAGAGAGAGTG 1361  
 Db 1800 sGluValProMetGluGlnValAlaGlnGluValHisValProSerGluPro----- 1818  
 QY 1362 GCTGGGAATCTATTAGGAATAGAAATGCTCGGTGGATCCATTCCTCCCTC 1421  
 Db 1819 -----SerProMetGluGluGluValHisValProSerProVal-----Proly 1834  
 QY 1422 CTTAAGACCA-----TCTGCAAGTCACACAGTGTCTGGGTAGACTTCTCGAG 1472  
 Db 1834 sGluGluProLysMetGluGluSerProGluGlnThrProThrProAspLeuIleSe 1854  
 QY 1473 ACAGAGAGAG-----GTTCCGGAACCTTTTCAACACATTACAAAGATTCTGGGA 1523  
 Db 1854 rAsnAsnGluSerGlnAspThrProGlyAlaValAsnAsnHisLeuHisGluAsnHisAs 1874  
 QY 1524 T-----AACAAACAT-----GTGAAAATGCTCTTG 1547  
 Db 1874 pAlaValGlnThrProLysLeuGlnProAlaSerGlnHisGlnValAlaGlnProSe 1894  
 QY 1548 TTTCGGAACAAACTTGTACCTGTGGAGATGAGATGCTGAGCGAATGTGCGAGGTAG 1607  
 Db 1894 rProArgProAlaValAlaProAspSerGlnGlnAsnGly----- 1907  
 QY 1608 GTGGGAGCTATTCAGCTGCTCTTCAACAAATTCACAGACCCAGAACTTG----- 1662  
 Db 1908 -----ProValLeuValSerGlnGlnSerGlnProSerProMetSerSe 1922  
 QY 1663 -AAGGATGCGATTCTGAATACATGCTGCAATATTCTAAGAATGGACTTTACAGCTTT 1721  
 Db 1922 rGlnGlnSerAspMetAlaGlnAsnLeuLeuLeuSerSerLys-----Asp11 1938  
 QY 1722 GGTGATTTCTGGATPAGTACTTGAAGAGCAGAGGCG----- 1761  
 Db 1938 eAsnAspLeuAlaAlaLysLeuHisLysAsnProGluAlaLeuAlaGlnAlaThrArgG1 1958  
 QY 1762 -----CAACATTATAT-----CAGTCCAT 1781  
 Db 1958 yAspCysSerGlyIlePheGlnHisLeuLeuHisAlaGlnGlyAsnGlyGlnAsnMe 1978  
 QY 1782 TTTACCTGATGCTGGAATTCACCTCTGCTGTCGCAATATTGTGACCCAGCAATACC 1841  
 Db 1978 tThrProGluMetLeuGlnLeuLysAlaAlaPhe----- 1989  
 QY 1842 ACTCTGAAACAGAGATGAATCATCTGTCAGATGTACAGAGACAGATCCAGTCT 1901  
 Db 1990 -PheAlaGlnGlnGlnGlnGlnAlaAsnGlnMetMetGln----- 2003  
 QY 1902 TTTAGCTAATGCTTTCTTGTGACATTTCCCGACGGAATGCCAAGATGAATCGGAGTA 1961  
 Db 2004 -----AlaLysMetLysGlnGln----- 2009  
 QY 1962 TTTAGTATCCACAGACATTAACTTCAATCGTTGTTTGAAGGAGCTTTCATCAAGGAACC 2021  
 Db 2010 -----ThrIleAsnLysAspArgIle-----LysGluG1 2019  
 QY 2022 AGAAAACTGAAACACTCTTCTGCTACTTTCGAGAGTGCACAGAGAAAAAACCCTACAGG 2081  
 Db 2019 nGluArgValLysArgMetTyrGluGluAsnGluArgLysValGlu----- 2034  
 QY 2082 ATTGGTGACATTTCACAGACAGATCTTGAAGATTTTCCAGATGCGAAAGGTGTA 2141  
 Db 2035 -----GluAspArgArgGluLysGlnArgLysGluG1 2045  
 QY 2142 GCCTCTGACACGCTTACACGCTCACTTACAGGGGTACCATAGAAGCGACCGCCAGCAT 2201  
 Db 2045 uGluArgLysArgLeu-AlaAlaAlaThrAlaAlaAlaThrMetAlaThrGlnLysAlaA 2065  
 QY 2202 GC-----TACAGGTGATTTTGCAA 2222  
 Db 2065 laGluAlaLeuLysGlnLysGlnGluValProArgHisGlyPheGlnHisValLeuSerM 2085  
 QY 2223 TCGTTTGTGGAGTGTGCTGCTGGGGACTTGTACAGAGAAATCAGATTTT 2282

Db 2085 etMetThrProGlu-----AlaArgSerLeuTyrGluGlnPheProGlyLeuS 2101  
 QY 2283 AATCAATCTGAAT 2296  
 Db 2101 erSerTyrIleAsn 2105  
 RESULT 13  
 A58198  
 N:serine/proline-rich FEL protein, splice form 1 - human  
 N:Alternate names: AF-4 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 02-Jul-1996 #sequence\_revision 25-Oct-1996 #text\_change 01-Dec-2000  
 C:Accession: A58198; 152572  
 R:Morrissey, J.J.M., 1993  
 Blood 81, 1124-1131, 1993  
 A:Title: A serine/proline-rich protein is fused to HRX in t(4;11) acute leukemias.  
 A:Reference number: 152572; PMID:8443374  
 A:Accession: A58198  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1213 <MO2>  
 A:Cross-references: GB:L25050; MID:9407323  
 A:Accession: 152572  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 'MAQS', 13-1213 <MO2>  
 A:Cross-references: GB:L25050; MID:9407323; PIDN:AAA36642.1; PID:9407324  
 C:Comment: This protein is one of several to form chimeric fusion proteins with the trit  
 C:Genetics:  
 A:Gene: GDB:MLJ2; AF-4; PEL  
 A:Cross-references: GDB:136792; OMIM:159557  
 A:Map position: 4q21-4q21  
 C:Keywords: alternative splicing  
 Alignment Scores:  
 Pred. No.: 0.00309 Length: 1213  
 Score: 162.00 Matches: 215  
 Percent Similarity: 33.21% Conservative: 137  
 Best Local Similarity: 20.28% Mismatches: 401  
 Query Match: 2.39% Indels: 307  
 DB: 2 Gaps: 48  
 US-09-302-812-5 (1-3814) x A58198 (1-1213)  
 QY 15 TCGGG-----TCCAGCATGATCGGGCCCGCTG-----GCA 50  
 Db 158 CysGlyProProAspSerGlnHisLeuThrGlnAspArgLeuGlyGlnGluGlyPheGly 177  
 QY 51 GCCTGTGACGAAAGCGCTGCGGGCGCTGCTGCGCGGAACTTCTGCGCGGACTGCTCGGACTC 110  
 Db 178 SerSerHisHisLysLysGlyAspArgArg-AlaAspGlyAspHisCysAlaSerValTh 197  
 QY 111 CCGAGCTTCTCGGAGCAGCAGGCGTCTTTCGACCCCAAGGAGCGTCCCGTCCAGTT 170  
 Db 197 rAspSerAlaProGluArgLysLeuSerProLeuSerProLeuSerPro----- 214  
 QY 171 CAGGCTTCTCGTCTCGGAGCTGCTGCTGCGGGCGGGGACCGACGACAGAGCAA 230  
 Db 215 ----ValProProLeuSerProIleHisSerAsnGlnGlnThrLeuProArgThrGlnG1 233  
 QY 231 CGCCTGCTGTTGTTTTCACAAAGAGCTATTACTTCTGGATGATTAAGAGACC 290  
 Db 233 ySerSerLysValHisGlySerSerAsnAsnSerLysGlyTyrCysProAlaLysSerPr 253  
 QY 291 CAAG-----ACAGCTGAATCAGAAAGTAAAGAAACAATACAGATTGCTCCAT 344  
 Db 253 oLysAspLeuAlaValLysValHisAspLysGlu-----ThrProGlnAspSerLe 270  
 QY 345 GATGAGTCTGTGCGAAGAGATAAC-----TTTACCCATTAAGGT 386  
 Db 270 uValAlaProAlaGlnProProSerGlnThrPheProProProSerLeuProSerLysSe 290

387 GGAATAATGGAAATGTTCTCAGCTAAATCTT-----GATAATC 428  
 Db : : : : :  
 290 rValalaMetGlnGlnProThrAlaTyValArgProMetAspGlyGlnAspGlnAl 310  
 QY : : : : :  
 429 ACCACAGAAAG-----AGTTACAGTATTGAAACCAACAGCAGAC 470  
 Db : : : : :  
 310 aProSerGluSerProGluLeuLysProLeuProGluAspTyL-----ArgGlnGlnTh 328  
 QY : : : : :  
 471 T-----CCGAGTGTGTCACAGTGCAGATGA 497  
 Db : : : : :  
 328 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 348  
 Db : : : : :  
 498 AGGGAAGCATGCAGACAGCTTTTGGCAGTGAGCTCCCGGGGAGCTCCGCTACCAAA 557  
 QY : : : : :  
 348 oSerGlnSerValGluGlnThrTySerAsnGluValHisCysValGluGluLysLeuLys 368  
 Db : : : : :  
 558 GCAGCTTAGTAATGCTAACATTGGTCAGTCACCCACACTGATGACACAGTGCAGACAGA 617  
 QY : : : : :  
 368 sGluMetThrHisSer-----TipProProLeuThrAlaLysHisThrProSerTh 386  
 Db : : : : :  
 618 TCATGAAGAAGACAGACAGCAATCAGCAGTTTCTTACCTATAAACTGCAAAATACAAA 677  
 QY : : : : :  
 386 rAlaGluProSerLys-----PheProPheProThrLysAspSerGlnHisVa 402  
 Db : : : : :  
 678 GCCACAGTAGGAGATGGCGCAGCCAGAGCAACTGTAGTCAGTGGATCTCGCCAGTC 737  
 QY : : : : :  
 402 lSerSerValThrGlnAsnGlnLysGlnTyAsp-----ThrSe 415  
 Db : : : : :  
 738 TGTGAAAGACTGTACAGGCTGTCAACAGGAGGAGTGGATGCTACACAGAGT---CC 794  
 QY : : : : :  
 415 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuGluAspAspLeuGl 435  
 Db : : : : :  
 795 TTTGTCAGATGTTGGTCGCGAGCAGCAATTCGAATCGACCAAAAATGACAAATTTGAC 854  
 QY : : : : :  
 435 nLeuSerAsp-----SerGluAsp-----SerAspSerGluGlnTh 447  
 Db : : : : :  
 855 TGGCAAGAAGACAGCCTAGTGTGCTCGCTCCATTGTGAGAAGAAAGTGCAGCTGAGTC 914  
 QY : : : : :  
 447 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu--- 466  
 Db : : : : :  
 915 ACCAATGATGTAGACAACTCGAAGAACAAGTGTCAAGATTTCAGACGATGAGAAAC 974  
 QY : : : : :  
 467 -ProValAlaSerAlaHisSerSerAlaGluSerGluSerThrSerAspSerAspSe 486  
 Db : : : : :  
 975 AGTCCAGCTTTTCAGCAAGATGATCGTCTCCCAACAGCAAAATAAACTTCAAG 1034  
 QY : : : : :  
 486 rSerSerAspSerGluSerGluSerSerSerSerSerSerGluGluAsnGluProLeuGl 506  
 Db : : : : :  
 1035 TTGCCAAGCAAGAAGAGCTAGTGGCTCTTAGGAAACGG-----TATT 1079  
 QY : : : : :  
 506 uThrProAlaProGluProGluProProThrThrAsnLysTipGlnLeuAspAsnTrpLe 526  
 Db : : : : :  
 1080 GACTAAGGGAAGTCAA-----GTTAG 1100  
 QY : : : : :  
 526 uThrLysValSerGlnProAlaAlaProProGluGlyProArgSerThrGluProProAr 546  
 Db : : : : :  
 1101 ATTGCATTTCCATTGAAGCAAGAAATATCTGGGACCGAGTCTTAATGCCAAG-- 1158  
 QY : : : : :  
 546 gaRgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 566  
 Db : : : : :  
 1159 -----CCATCTGGAACCTCTTCAGCTT----- 1182  
 QY : : : : :  
 566 rLysAspProProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 586  
 Db : : : : :  
 1183 -----AATGTAGAGTGTAGAGTTCC----- 1203  
 QY : : : : :  
 586 sProGlyLysArgSerCysGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 606  
 Db : : : : :  
 1204 -----AAGCAG----- 1209  
 QY : : : : :  
 606 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 626  
 Db : : : : :  
 1210 -----CATGGAAGAAAGGATTCTAAATTTACAGATCAITTTTCATGAGAATTTTCCAAAGTC 1262  
 QY : : : : :

626 rSerLeuGlnGlyGlyArgGluProGlyLeuLeuProTyGlySerArgAspGlnThrSe 646  
 Db : : : : :  
 1263 AGAGCAGAGAAAGAAACAATGTGAAGTCAGACATCAAAGAACAGAAAGAGATTC 1322  
 QY : : : : :  
 646 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaAlaSerAsnGluPr 666  
 Db : : : : :  
 1323 AAAATACATCCACCTAACCTCCCTCCAGAGAAAGTGGTGGAACTCTCTATTGAGGA 1382  
 QY : : : : :  
 666 Olys-----ProAlaValProProSerSerGlu----- 675  
 Db : : : : :  
 1383 AATGAGAAAAAGCTCGTGTGGATCCATTTGCCTTCTTAACCATCTGCAAGTCA 1442  
 QY : : : : :  
 676 ---LysLysLysHisLysSerSerLeuProAlaProSer----- 687  
 Db : : : : :  
 1443 CACAGTACTGTTCCGGTAGACCTTCTCAGAGCAGGAGGTTCCGAAACCTTTTCCAAC 1502  
 QY : : : : :  
 688 -----LysAlaLeuSerGlyProGluProAlaLysAs 698  
 Db : : : : :  
 1503 ACATTACAAAGATTGTGGGATAACAAACATGTGAAAATGCTTGTTCGGAACAAACTT 1562  
 QY : : : : :  
 698 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 717  
 Db : : : : :  
 1563 GTACCTGTGGAAGATGAGATGTGAGCAACTCAGGAGTGTGGAGCTCATTTCA 1622  
 QY : : : : :  
 718 -GlyProProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValVa 737  
 Db : : : : :  
 1623 GACTGCATCTTCTCAACAAATTCACACAGACCCAGAACTTGAAGGATCGCAATCTGAAATA 1682  
 QY : : : : :  
 737 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----- 753  
 Db : : : : :  
 1683 CAATGTGGCATATTTCAAGAAATGGGACTTTACAGCTTGTGTGATTTCTGGGATAAGT 1742  
 QY : : : : :  
 754 -----LysLe 755  
 Db : : : : :  
 1743 ACTTGAAGACAGAGGCCCAACATTTATATCATCTCAGTCCATTTTACCTGACATGTTGAAAT 1802  
 QY : : : : :  
 755 uLeuSerProLeuArgAspThrProProGlnSerLeu-----MetValLysIl 772  
 Db : : : : :  
 1803 TGCACCTCTCTGCCAAATATTTGCACCCAGCCA----- 1836  
 QY : : : : :  
 772 eThrLeuAspLeuLeuSerArgIleProGlnProProGlyLysGlySerArgGlnArgLy 792  
 Db : : : : :  
 1836 ----- 1836  
 QY : : : : :  
 792 sAlaGluAspLysGlnProProAlaGlyLysHisSerSerGluLysArgSerSerAs 812  
 Db : : : : :  
 1837 -----ATACCACCTCTCGAAACAGAGATG-----AATCA 1865  
 QY : : : : :  
 812 pSerSerSerLysLeuAlaLysLysArgLysGlyGluAlaGluArgAspCysAspAsnLy 832  
 Db : : : : :  
 1866 TTTCTGTACAGATG-----TCACAGAACAGATCGCCAGTCTTTTAGCTAA 1910  
 QY : : : : :  
 832 sLysIleArgLeuGluLysGluLysLysSerGlnSerSerSerSerSerHisLy 852  
 Db : : : : :  
 1911 TGTCTTCTTCGACATTTCCCGACGGAATGCCAAGATCAAAATCGAGATTTCTAGTTA 1970  
 QY : : : : :  
 852 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 872  
 Db : : : : :  
 1971 CCAGACANNTAACTTCAATCGGTGTTTGAAGACGTTTCATCAAGAAACCCAGAAATA-- 2028  
 QY : : : : :  
 872 uProProProProValSer-----SerSerSerGlnLysProAlaLysPr 887  
 Db : : : : :  
 2029 -----CTGAAACACTTCTCTGCTACTTTCAAGAGTCACAGAG-----AAAAAAC 2075  
 QY : : : : :  
 887 oAlaLeuLysArg-----SerArgArgGluAlaAspThrCysGlyGlnAspPr 903  
 Db : : : : :  
 2076 TACAGATTTGTCAGATTTTACAGACAGATCTTTGAGAT-----TTTCCAGATGGGA 2129  
 QY : : : : :  
 903 oProLysValProAlaValProArgValAsnHisLysAspSerSerLysProLysGlnAr 923  
 Db : : : : :  
 2130 AAGGTGTGAAGACCTCTGACAGCTTACACGCTTACAGCTTACAGAGTACCATA----- 2181  
 QY : : : : :



```

Db      367 IleAlaLeuPheTyrProAsnSerAsnAsnThrPheAspTyrThr-----SerAla 383
QY      1168 AACTCTTCT-----AGCTTAATGTAGAGTGTAGAGT 1200
Db      384 AsnGluAspPheLeuAlaThrArgAspPheLeuArgAsnLeuAspIleAsnVallysGlu 403
QY      1201 TCCAAGCAGCATGGAAGGATTCTAAATACAGATCATTTTCATGAGATTTCCCAAG 1260
Db      404 GluLysLysAsnIleGlnIleSerHisIleGlyIleAsnAspPheAsnTyrSerLeu 423
QY      1261 TCAGAGCAGACAGAAAGACAAATGTGAGTCAG-----ACATCA 1301
Db      424 IleGluLysGluSerAsnGluAsnSerLysThrPheGluPheProGlnGluLeuAsp 443
QY      1302 AAGAACAGAAAGAGATTCACAAATATACATCCACCTAACCTCCCTCCAGAGAGAAAGTG 1361
Db      444 LysAsnLysPheLysGlnAlaLysVallysPheThr---PheLysSerGlnSerGluGln 462
QY      1362 GCTGGAACTCTATTGAGGAATAGAAAATGCTGGTGGGATCCATTTCCCTTC 1421
Db      463 AlaTyrAsnSerLysArgSerValAsnLeuAspValHisIleAspAspLysPheAlaPhe 482
QY      1422 CTTAGACCATCTGCAAGTCACACAGTGAAGTCTCGGGTAGACCTCTGAGAGCAGGAGA 1481
Db      483 ThrLysThrIleGluTrpThrLysSerLeuGluSerIleTyrAsnGlnPheGluLys 502
QY      1482 GGTTCGGAACCTTTTCCAAACATTAACAAATTTGGGATAACAAACATGTGAAAT 1541
Db      503 ArgGluGlnArgTyr---AsnGluLeu-LysAspLeuIleLysLysTyrPheGlu-- 520
QY      1542 GCCTTGTGGGAACAAATCTGACCTGTGGAAGATGAGATGTGAGCGAAGTGCAGG 1601
Db      521 -----AsnArgLysAsnThrLeuProIleGluLeuProAsnTyrLysLysGlyLeuG 538
QY      1602 GAGTAGTGGGAGCTCATTCAGATGCATCTTCAACAAATTCACAGACCCACAGACTT 1661
Db      538 yser-----GluAsnLysGluLeuAlaLysSerThrLys---IleSerIle 552
QY      1662 GAAGGAT-----GCGATTCTGAATACAAT 1686
Db      552 uLysAspGluIleAspAlaThrPheLysGlyTyrGlyIleLeuLysTyrAsn 569

RESULT 15
A47328
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: A47328
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; PMID:93133824; PMID:8421688
A:Accession: A47328
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1403 <AND>
A:Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252
A:Experimental source: NK killer cells from adult blood
A>Note: sequence extracted from NCBI backbone (NCBI:122798, NCBI:122800)
C:Genetics:
A:Gene: GDB:NKTR
A:Cross-references: GDB:137171; OMIM:161565
A:Map position: 3p23-3p21
C:Map family: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

Alignment Scores:
Pred. No.: 0.00406 Length: 1403
Score: 160.50 Matches: 143
Percent Similarity: 33.47% Conservative: 102
Best Local Similarity: 19.54% Mismatches: 281

```

```

Query Match: 2.37% Indels: 206
DB: 1 Gaps: 30
US-09-302-812-5 (1-3814) x A47328 (1-1403)
QY      85 ACTTCTGGCGCGACTGCTCGGACTCCCGGAGACTTCCCTGGCAGCAGGCGGTGTTCTC 144
Db      591 ThrAlaSerLysSerSerHisSerArgSer-----
QY      145 GACCCCAAGGACGCTCCCGTCCAGTTCAGGTTCCTCGTCCCTCGTCCCGAGCTGCTCG 204
Db      602 -----ArgSerLysSerArgSerSer 608
QY      205 GGGCGGGGGGACCCGACACAGAGGCAACGCCACCTCGTTGTTTCAACAAAAGACTATT 264
Db      609 SerLysSerGly---HisArgLysArgAla----- 617
QY      265 ACTACTTGATGATACTTAAAGGACCCAGACAGCTGAATCAGAAAGTAAAGAAAACAAAC 324
Db      618 -----SerLysSerProArgLysThrAlaSerGlnLeuSerGluAsnLys 632
QY      325 AATCAAGAATTGACTCCATGATGAGTCTGTG---CAGAAAGATAACTTTTACCACAT 381
Db      633 ProValLysThrGluProLeuArgAlaThrMetAlaGlnAsnGluAsnValValGln 652
QY      382 AAGTGGGAAAATTTGGAAAATTTCTCAGCTAAATCTTGTAAATCAACC----- 432
Db      653 ProVal---ValAlaGluAsnIleProValIleProLeuSerAspSerProProSer 671
QY      433 -----ACA 435
Db      672 ArgTrpLysProGlyGluLysProTrpLysProSerTyrGluArgIleGlnGluMetLys 691
QY      436 GAAAGAGCTTCAAGTATTGAAACCAACAGCAGACTGCGAGTGTGTGAGTGGCAGAT 495
Db      692 AlalysThrThrHisLeuLeuProIleGlnSerThrTyrSerLeuAlaAsnIleLysGlu 711
QY      496 GAAGGGAAGCATGCAAGACAGCTTTTGGCAAGTGAAGCTCCCGCGGGAGCTCCGCTACCA 555
Db      712 ThrGly-----SerSerSerSerTyrHis 719
QY      556 AAGCAGCTTAGTAATGCTAATGCTAGTCAGTCAACCCACACT-----GATGACCACT 609
Db      720 LysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSerAspArgSer 739
QY      610 GACACAGATCATGAAGACAGACAGAC-----AATCAGCAG 645
Db      740 SerGluSerSerProArgSerArgSerArgSerArgSerArgSerArgSerArgSer 759
QY      646 TTCTTACACCTATAAACTTCAATACAAACCAACAGTAGGAGATGGCAGGCGACA 705
Db      760 TyrThrArgSerArgSerLeuAlaSerSer-----HisSerArgSerArg 774
QY      706 AGCAACTGTAGTGCAGTGATCTCCGAGTCTGTGAAAGACTGTACAGGTGTGTCACAG 765
Db      775 SerProSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArg 794
QY      766 GAGGAGGTGATGTCTCACCAGAGTCTTTGTCAGATGTTGTGTCGCGAGCAGCATTTGA 825
Db      795 -----SerSerSerTyrThrSerIleSerSerAspAsp---Gly 806
QY      826 ACTGGACCAAAAATGACACAAATTTGACTGCAAGAAAGACGCTAGGTGATTCGCT 885
Db      807 ArgArgAlaLysArgArgLeuArgSerSerGlyLysLysAsnSerValSerHisLysLys 826
QY      886 ---CCATTGTGAAAAGAAAGTGAAGCTGAGTCCCAATGCAATGATGTACAACTCGAGAAC 942
Db      827 HisSerSerSerSerGluLysThrLeuHisSerLysTyrValLysGlyArgAspArgSer 846
QY      943 AGTTGTCAAGATTCAGAGCAGATGAAGAAACAGTCCAGTCTTTGATGACGACAGATGAT 1002
Db      847 SerCysValArgLysTyrSerGluSerArgSerSerLeuAspTyrSerSerAspSerGlu 866

```



```

Db      1185  IleCyThr---ProAspArgSerProAlaLysVaIgluThrSerProLeugly 1203
              |||||  |||  |||:||||:  |||  |||  |||  :|||
QY      1969  -----TACCCAGACATTAAAC 1983
              |||||  |||||  |||||  |||||  |||||
Db      1204  AsnAlaArgLeuAspThrProAspIleAsn 1213
              |||||  |||||  |||||  |||||  |||||

Search completed: May 26, 2004, 17:19:14
Job time : 188.361 secs

```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:32:43 ; Search time 31.9083 Seconds  
(without alignments)  
12447.891 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 9999gactgtgtgtgcgggg.....aatcattgtcagaataaaa 3814

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delopt 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xl  
-Q/cgn2.1/USFTO spool\_p/6333148/runat 26052004.150053.5664/app.query.fasta\_1.12437  
-DB=SwissProt 42 -OPMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148 @CGN 1 1.128 @runat 26052004.150053.5664 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	176	2.6	6632	UN89 CAEEL	Q01761 caenorhabdi
2	172	2.5	1647	Q01124 CAEEL	Q10124 caenorhabdi
3	171	2.5	997	BIRI_SCHPO	Q14064 schizosacch
4	166.5	2.5	699	NP14 HUMAN	Q14978 homo sapien
5	166	2.5	1210	AF4 HUMAN	P51825 homo sapien
6	165	2.4	2314	AKA6 RAT	Q9wvc7 rattus norv
7	162	2.4	5038	PCLO_MOUSE	Q9y6v0 mus musculu
8	161	2.4	5147	PCLO_HUMAN	Q30414 homo sapien
9	160.5	2.4	1462	NKCR_HUMAN	P34414 homo sapien
10	159	2.3	687	DSPP RAT	Q62598 rattus norv
11	158	2.3	5085	PCLO RAT	Q9jke6 rattus norv
12	157.5	2.3	559	ENL_HUMAN	Q03111 homo sapien
13	157.5	2.3	934	DSPE_MOUSE	P97399 mus musculu
14	157.5	2.3	1359	ATRY_CAEEL	Q9u760 caenorhabdi
15	157.5	2.3	2211	FA5_BOVIN	Q28107 bos taurus
16	156	2.3	1274	ENAM_MOUSE	O55196 mus musculu
17	155.5	2.3	917	TBC2_HUMAN	Q9y6x2 homo sapien
18	155	2.3	1489	YGPO_YEAST	P53115 saccharomyc

19	152.5	2.3	406	1	SR40 YEAST	P32583 saccharomyc
20	152.5	2.3	5120	1	PCLO_CHICK	Q9pu36 gallus gall
21	152	2.2	1983	1	TF20_MOUSE	Q9epq8 mus musculu
22	151	2.2	913	1	IF38_HUMAN	Q96613 homo sapien
23	151	2.2	1093	1	YKDS_CAEEL	Q03563 caenorhabdi
24	151	2.2	1447	1	BUDA YEAST	P47136 saccharomyc
25	150.5	2.2	513	1	DMP1_HUMAN	Q13316 homo sapien
26	150	2.2	489	1	DMP1_RAT	P98193 rattus norv
27	150	2.2	1070	1	PVDR_PLAYS	P22290 plaesmodium
28	150	2.2	1337	1	DEXT_STRDO	P39653 streptococc
29	149.5	2.2	2128	1	SPCB_MOUSE	P15508 mus musculu
30	149.5	2.2	2137	1	SPCB_HUMAN	P11277 homo sapien
31	149.5	2.2	5596	1	MDN1_HUMAN	Q9nu22 homo sapien
32	149	2.2	1453	1	NKCR_MOUSE	P30415 mus musculu
33	148	2.2	761	1	Y218_HUMAN	Q93075 homo sapien
34	148	2.2	1358	1	SIR4_YEAST	P11978 saccharomyc
35	148	2.2	1380	1	ZMS1_YEAST	P46974 saccharomyc
36	147.5	2.2	943	1	CENC_HUMAN	Q03188 homo sapien
37	147.5	2.2	1031	1	YDG9_SCHPO	Q10496 schizosacch
38	147.5	2.2	2738	1	PGCV_RAT	Q9etb4 rattus norv
39	147	2.2	823	1	UB16_HUMAN	Q9y5t5 homo sapien
40	146.5	2.2	3122	1	DPOZ_MOUSE	O61493 mus musculu
41	146	2.2	1395	1	SP41_YEAST	P38904 saccharomyc
42	146	2.2	3924	1	ANK2_HUMAN	Q01484 homo sapien
43	145.5	2.1	1282	1	BMS1_HUMAN	Q14692 homo sapien
44	145.5	2.1	2442	1	CBP_HUMAN	Q92793 homo sapien
45	145	2.1	574	1	MIG1_CANAL	Q9y7g2 candida alb

## ALIGNMENTS

RESULT 1  
UN89 CAEEL  
ID UN89 CAEEL STANDARD; PRT; 6632 AA.  
AC Q01761; Q17362;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
GN UNC-89 OR C09D1.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RC STRAIN=Bristol N2;  
RX MEDLINE=96180278; PubMed=8603916;  
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
assembly, encodes a giant modular protein composed of Ig and signal  
transduction domains.";  
RL J. Cell Biol. 132:835-848 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Du Z., Le T.T., Wilson R.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP REVISIONS.  
RL Waterston R.;  
CC Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -|- FUNCTION: Structural component of the muscle M-line. Myofilament  
lattice assembly begins with positional cues laid down in the  
basement membrane and muscle cell membrane. UNC-89 responds to  
these signals, localizes, and then participates in assembling an  
M-line.  
CC -|- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
CC -|- SIMILARITY: Contains 1 DAL-homology (DH) domain.  
CC -|- SIMILARITY: Contains 1 fibronectin type III domain.  
CC -|- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
CC -|- SIMILARITY: Contains 1 PH domain.  
CC -|- SIMILARITY: Contains 5 RCSD domains.



193 GCTCGCTCGGGGGGGGAGCGGAGCGCAGAGGC---AAGCCACCTCGTTGTT--- 246 QY  
1478 GluAspValLysSerProValLysLysGluLysSerProAspAlaThrAsnIleValGlu 1497 Db  
247 -----TTCAACAAGACATATTACTTGTGGATGATACT 282 QY  
1498 ValSerSerGluThrThrIleGluLysThrGluThrMetThrHis 1517 Db  
283 AAAGGACCAACAGACACTGAATCAGAAAGTAAGAAACAACAATACAGAAATTGACTCC 342 QY  
1518 GluSerGluGluSerArgThrSerValLysLysGluLysThrProGluLysValAspGlu 1537 Db  
343 ATGATGAGTTGTGTCAGAAAGATAACTTTTACCACATAAG-----GTGGAATAATGTG 396 QY  
1538 LysProLysSerProThrLysLysAspLysSerProGluLysSerIleThrGluGluIle 1557 Db  
397 GAAATGTTCTCAGCTAAATCTTGATAATCACCACACA-----CAAAAGAGTTCCACAG 450 QY  
1558 -----LysSerProValLysLysGluLysSerProGlu 1568 Db  
451 TATTTGAACCAACAGACAGACTGCGAGTGTGTGCAAGTGGCAGAAATGAAGGAGCATGCA 510 QY  
1569 LysValGluGluLysProAlaSerProThrLysLysGluLysSerProGluLysProAla 1588 Db  
511 GAACAGCTTTTGCAAGTGTGCTCCCGGGGAGTCCGCTA----- 552 QY  
1589 SerProThrLysLysSerGluAsnGluValLysSerProThrLysLysLysGluLysSerPro 1608 Db  
553 -----CCAAAGCAGCTTGAATGTAATGTAACATTGGT 582 QY  
1609 GluLysSerValValGluGluLysSerProLysGluLysSerProGluLysAlaAsp 1628 Db  
583 CAGTCAACCCACACT-----GAT 600 QY  
1629 AspLysProLysSerProThrLysLysGluLysSerProGluLysSerAlaThrGluAsp 1648 Db  
601 GACCACAGTGCACAGATCATATGAAGAA-----GACAGACAAATCAGCAGTTCTTACA 654 QY  
1649 ValLysSerProThrLysLysGluLysSerProGluLysValGluGluLysProThrSer 1668 Db  
655 CCTATAAATCTTCAATACAAAGCCACAGTAGGAGTGGCAGGCGCAGAACTGT 714 QY  
1669 ProThrLysLysGluSerSerProThrLysLysThrAspGluValLysSerProThr 1688 Db  
715 AAGTGCAGTGGATCTCGCCAGTCTGTGAAGACTGTACAGCTGTCAACAGGAGGAGTG 774 QY  
1689 LysLysGluLysSerProGluThrValGlu----- 1698 Db  
775 GATGTGCTACAGAGTCTTTGTGTGATGTTGGTCCGAGGACATTTGGAAGTGGACCA 834 QY  
1699 -----GluLysProAlaSerProThrLysLysGlu----- 1708 Db  
835 AAAAATGACAACAAATGACTGGACAGAA-----AGCAGCTAGTGTGATTCGCTCCATTT 891 QY  
1709 LysSerProGluLysSerValValGluValLysSerProLysGluLysSerProGlu 1728 Db  
892 GAGAAAGAAAGTGTGCTGAGTACCAATGATGATGATGATGATGATGATGATGATGATGAT 951 QY  
1729 LysAlaGluGluLysProLysSerPro-----ThrLysLysGluLysSerProGlu 1745 Db  
952 GATTGAGAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011 QY  
1746 LysSerAlaAlaGluGluValLysSerProThrLysLysGluLysSerProGluLysSer 1765 Db  
1012 -----CAACAGCAAAATAACTTTCAAGTTCGCAAGCAAGCAAGCAAGTGTGAT 1056 QY  
1766 AlaGluGluLysProLysSerProThrLysLysGluSerSerProValLysMetAlaAsp 1785 Db  
1057 GCGCATCTTAGAAACGCTATTTGACTAAGGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1116 QY  
1786 AspGluValLysSerProThrLysLysGluLysSerProGluLysValGlu----- 1802 Db

1117 GAAGGAGAAATAATGCTGGGACCACTGACTTAAATGCGCAAGCCATCTGGAACCTTCT 1176 QY  
1803 -----GluLysProAlaSerProThrLysLysGluLysThrProGluLysSerAlaAla 1820 Db  
1177 AGCCTTAATAGTAGTGTAGAGTTCCAGCAGCATGGAAAAGGATCTCTAAATTACA 1236 QY  
1821 Glu-----GluLysSerProThrLysLysGluLysSerProSerProThr 1837 Db  
1237 GATCATTTTCAGAGATTTCCAGTCAGAGCAGAGCAAGAAAACAATGT---GAAGTC 1293 QY  
1838 -----LysLysThrGlyAspGluSerLysGluLysSerProGluLys 1851 Db  
1294 AGACATCAAGAACAGAAAGAGATTCACAAATACATCCACCTAACTCCCTCCAGAG 1353 QY  
1852 ProGluGluLysProLysSerProThrProLysLysSerProProGlySerProLysLys 1871 Db  
1354 AAGAAGTGGCTGGAACTCTTATGAGGAAATGAGAAAATGCTCGGTGGATCCAT 1413 QY  
1872 LysLys-----SerLysSerProGluAlaGluLysProPro----- 1883 Db  
1414 TTGCTTCTCTTAAGACCATCTCGAAGTCACAGTCACTGTTGGGTAGACCTTCTGAGA 1473 QY  
1884 AlaProLysLeuThrArgAspLeuLysLeuGlnThrValAsnLysThrAspLeuAlaHis 1903 Db  
1474 GCAGGAGAGGTCCGAAACCTTTTCCACACACATTCACAAAGATTTGTGGATACAAACAT 1533 QY  
1904 Phe-----GluValValValGluHisAlaThrGluCysLysTrpPheLeuAspGlyLysGlu 1922 Db  
1534 GTGAAAATCCCTGTTGCGGAACAAACTGTTGACCTGTGAGAGATGAGATGTTGAGCGCA 1593 QY  
1923 Ile-----ThrThrAlaGlnGlyValThrValSerLysAspGlnPheGluPhe 1939 Db  
1594 ACTGCAGGAGTAGTGGGAGCTCATTGACAGTGCATCTTCAACAAATTCACACGACCC 1653 QY  
1940 ArgCysSer-----IleAspThrThrMetPheGlySerGlyThr----- 1952 Db  
1654 CAGAACTTCAAGGATGCGATCTCTGAAATACATGTGCGCATTTCT----- 1698 QY  
1953 -----ValSerValValAlaSerAsnAlaAlaGlySerValGluThrLysThr 1968 Db  
1699 -----AAGAAATGGGACTTTACA----- 1716 QY  
1969 GluLeuLysValLeuGluThrProLysGluThrLysLysProGluPheThrAspLysLeu 1988 Db  
1717 -----GCTTTGTTGATTTCTGGGATAAGTACTTGAAGAACAGCAGAGGCCCAACAT--- 1767 QY  
1989 ArgAspMetGluValThrLysGlyAspThrValGlnMetAspValIleAlaLeuHisSer 2008 Db  
1768 ---TTATATCAG-----TCC 1779 QY  
2009 ProLeuTyLysTrpTyLysGlnAsnGlyAsnLeuGluAspGlyLysAsnGlyValThr 2028 Db  
1780 ATTTTACCTGACATGTTGAAATTCGACTGTGCTGCAATATTTGACCCGCCCAATA 1839 QY  
2029 IleLysAsnGluGluAsnLysSerSerLeuIleIleProAsn-----AlaGlnAspSer 2046 Db  
1840 CCATCTCTGAAACAGAAAGTGAATCTCTGTCAGATGTCACAGGACAGATGCGCAGT 1899 QY  
2047 GlyLysIleThrValGluAlaSerAsnGluValGly---SerSerGluSerSerAlaGln 2065 Db  
1900 CTTTGTAGCTAATGCTTTCTTCTGACAC-----TTTCCCGCA----- 1935 QY  
2066 LeuThrValAsnProProSerThrThrProIleValValAspGlyProLysSerValThr 2085 Db  
1936 -----CGAATGCAAGATGAATCGAGTATTTCTAGTATCCCA----- 1974 QY  
2086 IleLysGluThrGluThrAlaGluPheLysAlaThrIleSerGlyPheProAlaProThr 2105 Db  
1975 -----GACATTAATCTCAATCGGTTCTTTGAAGGACGTTTCATCAAGGAAACCAAGAA 2025 QY  
2106 ValLysTrpThrIleAsn---GluLysIleValGlu-----GluSerArgThrIleThr 2122 Db  
2026 AAACCTGAAACACTCTTCTGCTACTTTTCGAAGAGTCACAGAGAAAAA-----CCTACA 2079 QY

```

Db      2123 ThrIleLysThrGluAspValTyrThrLeuLysIleSerAsnAlaLysIleGluGlnThr 2142
QY      2080 GGATTGGTGACATTACACAGCAGT-----CTTGAAGATT 2118
Db      2143 GlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAlaAspLeu 2162
QY      2119 CCAGATGGGAAAGGTGTGAAAGCCTCTGACACGCTTACACGCTCACTTACGAGGGTACC 2178
Db      2163 LysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThrAspLysValAla 2182
QY      2179 ATGAAGCGAACGGC---CGAGGCATCTACAGTGGATTTCAMATCGTTTGTGGGA 2235
Db      2183 AspGluGlyGluProLeuArgTrpAsnLeuGluLeuAsp----- 2195
QY      2236 GGTGGTGTGACTGTGGCGGACTGTGTACAGAGAAATCAGATTTTAAATCAATCCTGAA 2295
Db      2196 GlyProSerProGlyThr-----GluValSerTrpLeuLeuAsnGlyGln 2210
QY      2296 TTGATTGTTTCACGGCTGTTCACTGAGTGTGTCATCACATGAG---TCTCTATTATTC 2352
Db      2211 ProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGlyThrTyrHisVal 2230
QY      2353 ACAGGTACTGAACAGTACAGTACATACAGGC 2385
Db      2231 ThrIleAlaGluAlaLysProGluMetSerGly 2241

RESULT 2
POEI_CABEL
ID POEI_CABEL STANDARD; PRT; 1647 AA.
AC Q10124; Q8MQ26; Q8MQ27; Q8MQ28; Q8MQ29; Q8MQ30; Q8MQ31;
DT 01-FEB-1996 (Rel. 33, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE PolyQ enhancer protein 1.
GN POE-1 OR F52C9.8/F52C9.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C), FUNCTION, AND SUBCELLULAR
  LOCATION.
RX MEDLINE=22388320; PubMed=12486229;
RA Faber P.W., Voisine C., King D.C., Bates E.A., Hart A.C.;
RT "Glutamine/proline-rich POE-1 proteins protect Caenorhabditis elegans
  neurons from huntingtin polyglutamine neurotoxicity.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17131-17136(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favell A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RA Waterston R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Protects neurons from the toxic effects of expanded
  poly-Q disease proteins. It is not known if this is done via
  participation in the pathogenic mechanism underlying poly-Q-
  induced neurodegeneration or acting as a genetic modifier of the
  age of onset or progression of neurodegeneration.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
  Event=Alternative splicing; Named isoforms=6;
  Name=b; Synonyms=Pqe-1A;
  IsoId=Q10124-1; Sequence=Displayed;
  Name=a; Synonyms=Pqe-1C;
  IsoId=Q10124-2; Sequence=VSP_008069, VSP_008070;
  Name=c; Synonyms=Pqe-1B;
  IsoId=Q10124-3; Sequence=VSP_008068;
  Name=d;
  IsoId=Q10124-4; Sequence=VSP_008065;

```

```

CC      Note=No experimental confirmation available;
CC      Name=e;
CC      IsoId=Q10124-5; Sequence=VSP_008063, VSP_008064;
CC      Note=No experimental confirmation available;
CC      Name=f;
CC      IsoId=Q10124-6; Sequence=VSP_008066, VSP_008067;
CC      Note=No experimental confirmation available;
CC      -1- DOMAIN: The Gln/Pro-rich N-terminus and the Arg/Asp/Glu/Lys-rich
  charged domain are critical in protecting glutamatergic ASH
  sensory neurons from degeneration. ASH neurons expressing isoforms
  lacking these domains show progressive degeneration.
CC      -1- SIMILARITY: To mammalian protein GOR, yeast YGR276C and X.laevis
  XPMC2.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC      EMBL; U39850; AAM45366.1; -
CC      EMBL; U39850; AAM45367.2; -
CC      EMBL; U39850; AAM45368.1; -
CC      EMBL; U39850; AAM45369.1; -
CC      EMBL; U39850; AAM45370.1; -
CC      EMBL; U39850; AAM45371.1; -
CC      PIR; T16421; T16421.
CC      WormPep; F52C9.8a; CE30802.
CC      WormPep; F52C9.8b; CE30998.
CC      WormPep; F52C9.8c; CE30672.
CC      WormPep; F52C9.8d; CE30673.
CC      WormPep; F52C9.8e; CE30674.
CC      WormPep; F52C9.8f; CE30675.
CC      InterPro; IPR006055; Exonuclease.
CC      Pfam; PF00929; Exonuclease; 1.
CC      SMART; SM00479; EXOIII; 1.
CC      Nuclear protein; Alternative
  splicing.
CC      GLN/PRO-RICH.
CC      ARG/ASP/GLU/LYS-RICH.
CC      EXONUCLEASE.
CC      Missing (in isoform e).
CC      /FTid=VSP_008063.
CC      DAKSENRAKEKQPMN -> MNVSSPSTQASHNQFQ
  (in isoform e).
CC      /FTid=VSP_008064.
CC      Missing (in isoform d).
CC      /FTid=VSP_008065.
CC      Missing (in isoform f).
CC      /FTid=VSP_008066.
CC      AKQKPAVAPVKILDFSVGRTFTAIRQTAKILVDTFLERD
  SPNAREAQEFELSIKQYTDGKYNIRINIGKVAARKENT
  SGILEVKNKA -> MSTYQSAPFLVHAHAYQPLSPSPSP
  PHGYLPPTSNPMWCSTSPLYFSPQWVQPLPQOHMYPS
  TSAGPGMYTDPNHYVQIQISQ (in isoform f).
CC      /FTid=VSP_008067.
CC      Missing (in isoform c).
CC      /FTid=VSP_008068.
CC      KQDAKSEN -> VRRCGGDNV (in isoform a).
CC      /FTid=VSP_008069.
CC      Missing (in isoform a).
CC      /FTid=VSP_008070.
CC      SQ SEQUENCE 1647 AA; 182261 MW; AJA66BA183BD650 CRC64;
  Alignment Scores:
  Pred. No.: 0.000527
  Score: 172.00
  Percent Similarity: 31.27%
  Best Local Similarity: 18.73%
  Query Match: 2.54%
  DB: 1
  Length: 1647
  Matches: 194
  Conservative: 130
  Mismatches: 367
  Indels: 345
  Gaps: 43

```



QY 1690 -----GCATATCTTAAGAAATGGGACTTT----- 1713  
 Db 1207 ProAlaLysGlnLysProAlaValProLysIleLeuAspPheSerValGly 1226  
 QY 1714 -----ACAGCTTTGGTTGATTTCTGGGATAAGGTACTTGAA 1749  
 Db 1227 ArgThrPheThrAlaIleArgGlnThrAlaIleLysLeuValPheAspThrPheLeuGlu 1246  
 QY 1750 GAA-----GCAGAGCCCAACATTTATATCATCATGTCATTTACCTGAC 1791  
 Db 1247 ArgAspSerProAsnAlaAlaArgGluAlaGlnGluPheGluLeuSerIleAlaLysGln 1266  
 QY 1792 ATGGTGAATTCACCTCTGCTGCCAAATATTGACCCAGCAATACCACTCCCTGAAA 1851  
 Db 1267 TyrThrAspGlyGlnLysTyrArgIleAsnIleGlyHisLysValAlaAlaLeuArgLys 1286  
 QY 1852 CAGAAGATGATCATCTCTGCACCATGCACAGAACAGATCGCC-----AGTCTTTTA 1905  
 Db 1287 GluAsnThrSerGlyIleLeuGluValAsnLysAsnAlaValSerHisAspLysIleLeu 1306  
 QY 1906 GCTAATGCTTTCTCTGCACATTTCCCGCAGGAAATCCAGATGAATCGGATATCT 1965  
 Db 1307 AlaGlyGly-----ProLysAspAsnCysThrValAlaAlaArgGlyArgLys 1321  
 QY 1966 AGTTACCCAGACATTAATCTCAATCGGTGTTTGAAGACGTTTCATCAAGAAACCAAGAA 2025  
 Db 1322 ThrHisValAsp-----HisArgGlnLeuSerIleGlu 1332  
 QY 2026 AAATCGAAACACTCTTCTGCTACTTTTCGAGAGTCACAGAGAAAACCTCAGGATG 2085  
 Db 1333 LysLeuHisProLeuLeuGlnPheLysLeuThrThrSerGluLeuGluThrAsnAla 1352  
 QY 2086 GTGACATTTACAAGACAGAT----- 2106  
 Db 1353 TyrProMetArgArgAspGlySerThrLysAlaValSerIleAlaAspThrValTyrThr 1372  
 QY 2107 -----CTTGAAGATTTCCAGATCGGAAAGGTGTGAAAGCCTCTG 2148  
 Db 1373 GlnAsnLysLysMetPheLeuAspAspTyrAspMetSerArgAsnCysSerArgCysAsn 1392  
 QY 2149 ACACGCTTACAGCTCACTACGAGGTACACATAGAGCAACGCCGAGGATCTCTACAG 2208  
 Db 1393 LysGluPheLysLeuSerProAsnGlyThr-----MetIlearg 1405  
 QY 2209 GTGGAATTTGCAATCGTTTGTGGAGGTGGTGCTGAGTGTG-----CGGACTT 2259  
 Db 1406 SerThrGlyIleCysArgTyrHisAsnArgGlyValAla-IleAsnGlyLysArgAspTh 1425  
 QY 2260 GTACAAGAGAAATCAGATTTTATCAATCTCAATGATTTGTTTCAC-----GG 2310  
 Db 1425 rPheArgLysArgTyr-----SerCysAsnGluGluPheAsnValAlaLeuGlu 1442  
 QY 2311 CTGTTCTCAGGTGCTGTCATCAATGAGTCTCTATTATFACACAGGTGATGAACATAC 2370  
 Db 1442 yCysLysPheSerAspValHisValThrAspGlnLeuPheLysLysGluLeuSerThrPh 1462  
 QY 2371 AGTGAATACACAGGCT----- 2386  
 Db 1462 eValSerThrProValProValProAsnAspGlnArgSerThrArgValTyrAlaLeuAs 1482  
 QY 2387 -----ATGCTGAACCTTATCGTTGGCCCGCAAGCCATGAAGATG 2425  
 Db 1482 pCysGluMetValTyrThrIleAlaGlyProAlaLeuAlaArgLeu 1497

RESULT 3

BIR1 SCHPO

ID BIR1 SCHPO STANDARD; PRT; 997 AA.

AC 014064; QUSG4;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Birl protein (Chromosome segregation protein cut17).

GN BIR1 OR CUT17 OR PBH1 OR SPCC962.02C OR SPCC31810.10C.

OS Schizosaccharomyces pombe (fission yeast)  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 [1]  
 RN SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RP MEDLINE=21439264; PubMed=11554922;  
 RX Morishita J., Matsusaka T., Goshima G., Nakamura T., Tatebe H.,  
 RA Yanagida M.;  
 RT "Bir1/Cut17 moving from chromosome to spindle upon the loss of  
 RT cohesion is required for condensation, spindle elongation and  
 RL repair."; Genes Cells 6:743-763 (2001).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RC MEDLINE=21848401; PubMed=11859360;  
 RX Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skellton J., Simmonds R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Wainman F., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Woodward J., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Mostl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Wambutt R., Purnelle B.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.B., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."; Nature 415:871-880 (2002).  
 [3]  
 RN CHARACTERIZATION.  
 RP MEDLINE=99398681; PubMed=10468581;  
 RX Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,  
 RA Vaux D.L., Lithgow T.;  
 RT "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell  
 RT division."; Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175 (1999).  
 [4]  
 RN CHARACTERIZATION.  
 RP MEDLINE=21850422; PubMed=11861551;  
 RX Rajagopalan S., Balasubramanian M.K.;  
 RA "Schizosaccharomyces pombe Bir1p, a nuclear protein that localizes to  
 RT kinetochores and the spindle midzone, is essential for chromosome  
 RT condensation and spindle elongation during mitosis."; Genetics 160:445-456 (2002).  
 [5]  
 RN FUNCTION.  
 RP MEDLINE=20035862; PubMed=10571085;  
 RX Rajagopalan S., Balasubramanian M.K.;  
 RA "S. pombe Pbp1p: an inhibitor of apoptosis domain containing protein  
 RT is essential for chromosome segregation."; FEBS Lett. 460:187-190 (1999).  
 CC -!- FUNCTION: Seems to act in the pleiotropic control of cell  
 CC division. Has a role in chromosome segregation by recruiting  
 CC condensin and ark1 kinase to appropriate sites as the cell  
 CC progresses through mitosis.  
 CC -!- SUBCELLULAR LOCATION: Nuclear. Interacts with the outer  
 CC centromeric regions of the chromosomes during interphase. After

CC chromatid separation moves to the middle of the spindle.

CC -!- SIMILARITY: Contains 2 BIR repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB031034; BAA83415.1; -;

DR EMBL; AL031323; CAA20434.1; -;

DR EMBL; AL121859; CAB58376.1; -;

DR PIR; T43523; T43523.

DR HSP; O13490; 1QBH.

DR GeneDB\_SPombe; SPCC962.02c; -;

DR InterPro; IPR001370; BIR.

DR Pfam; PF00653; BIR; 2.

DR SMART; SM00238; BIR; 2.

DR PROSITE; PS01282; BIR REPEAT 1; FALSE\_NEG.

DR PROSITE; PS0143; BIR REPEAT 2; 2.

KW Cell division; Mitosis; Nuclear protein; Repeat.

FT REPEAT 25 99 BIR 1.

FT REPEAT 120 194 BIR 2.

FT DOMAIN 80 83 POLY-ASP.

FT DOMAIN 312 319 POLY-ASP.

FT DOMAIN 487 490 POLY-SER.

SQ SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489F4 CRC64;

# Alignment Scores:

Pred. No.: 0.000527 Length: 997

Score: 171.00 Matches: 188

Percent Similarity: 34.44% Conservative: 133

Best Local Similarity: 20.17% Mismatches: 344

Query Match: 2.52% Indels: 267

DB: 1 Gaps: 45

US-09-302-812-5 (1-3814) x BIR1\_SCHPO (1-997)

QY 46 TGGGAGCCC-----TGACGAAA 63

DB 173 TrpGluProAspAspProTyrThrGluHisLysArgArgAlaaspCysValPhe 192

QY 64 GCGCGCTGGGCGCGCTGGAACTTCTGCGCGAGCTGCTCGGACTCCCGGAGCTTCCCT 123

DB 193 PheThrTrpLysAspProAsnSerLeuSerProThr-----LysLeuSerPheLeu 209

QY 124 GCGAGGAGAGCGGTTCGACCCCAAGGAGCGCTCCCGTCAGTTCAGGTCCTCCG 183

DB 210 SerThrSerAsn-----IleAspProGluAspLeuThrGluAspAsnSerIleLeuPro 227

QY 184 TCCTCGCCAGCGCTGCTCTCGGGGCGGCGGACCGCACAGAGGCAAGCCACCTCGTTT 243

DB 228 ValSerPro----- 230

QY 244 GTTTTCAACAAAGACTATTACTTCTGGTGGATGATAAAGAGCCCAAGACAGCTGAA 303

DB 231 -----ThrArgAspSerThrLysSerHisLysThrLeuAsn 242

QY 304 ----TCGAAAGTAAGAAACACAC-----AATACAGAAATTGATCTCCATGATG- 348

DB 243 PheSerProSerArgLysAsnAsnLeuAsnAlaArgProLeuThrMetSerLeuYrThr 262

QY 349 ---AGTTCTGTGCAGAAAGACTTTTACCCACATAAGGTGGAAAAATTGAAAAATGTT 405

DB 263 AsnThrSerGluLysAspSer-----GlnProThrArgAlaProGlnSerProThrLys 281

QY 406 CCTCAGCTA-----AATCTTGATAAATCAACCCAGAGAAAGAGTTCCACAG 450

DB 282 ProValLeuLeuThrAlaProArgArgLysAsnLysSerPro----- 295

QY 451 TATTTGAACCAACAGCAGACTGCGAGTGTGTGCAG-----TGGCAGAAT 495

DB 296 -----LysLysSerLysProAlaValPheLysProValLysProLlePheSerAsp 312

QY 496 GAAGGAAGCATGCAGAACAGCTTTTGGCAAGTGAAGCTCCCGCGGGACTCCGCTACCA 555

DB 313 GluAspGluAspAspAspLeuThrAlaSerGln-----ProPheSer 327

QY 556 AAGCAGCTTGAATAT-----GCTAACATTTGGTGTAGTCAACCCACATGATGACACAGT 609

DB 328 LysGlyIleCysAsnAspSerMetGlnValAlaLysLysAsnPheThrGluGluIlePro 347

QY 610 GACACAGATCATGAGAGACAGACAAATCAGCAGTTTCTTACACCTATAAACTTGA 669

DB 348 LeuLysGluAspGluLysAspAsnGlnLeuGluHisLeuValSerProAlaThrSerVal 367

QY 670 AATCAAAAGCCAAACAGTAGGAGATGGCGAGGCGCAGAACCACTGTAAGTGCAGTGGATCT 729

DB 368 HisThr----- 369

QY 730 CGCCAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTGTACCAGAG 789

DB 370 -----ThrValSerAspIleThrGlyHisGlnSer----- 379

QY 790 AGTCCTTTGTGTCAGATGTTGGTCCCGCAGCAGCATTGGAACTGGACCAAAAAATGACACAAA 849

DB 380 -----ValThrAspGluSerAspGlu----- 386

QY 850 TTGACTGGACCAAGAACGAGCCTAGTGTGCTCCATTTGAGAAAAGAGTGAAGCT 909

DB 387 -----GlnAsnAsnCysMetSerThrProProLysIleGluIleGluSerLysIle 403

QY 910 GAGTCACCAATGGATGTAGACAACTCGAGAACAGTTGTCAAGATTCAGAACGAGATGAA 969

DB 404 GluGluGluIleSerValValSer-----LysSerLysGluLysSerSer 418

QY 970 GAAACAAGTCCAGTCTTTGATGAGCAAGATGATCGTCTCTCCCAACACAGCAATAAAT 1029

DB 419 SerValSerSerValGlyLysGluGlnAsnHisThrGluLysGlnValAlaIleGluThr 438

QY 1030 TCAAGTCCCAAGCAAGAGAACTGATGCGCATCTTAGGAAACGGTATTGTGACTAAGGA 1089

DB 439 ProGluGlnGlnLysValGluLysGluAspGlu----- 449

QY 1090 AGTGAAGTTAGATTCATTTCCATTTTGAAGAGAA-----AATAATGCTGGGACCACT 1143

DB 450 -----HisLeuAsnLeuGlnGlySerPheIleGluGluSerThrLysGln 464

QY 1144 GACTTAAATGCCAAGCCATCTGGAACCTTCTAGCTTAAATGTAGAGTGAAGATTCCT 1203

DB 465 ProfileSerSerLysProSerThrSerSerProAspMetThrAspAlaIleThrGlyGly 484

QY 1204 AAGCAGCATGGAATA-----AGGATTCCT-----AAA 1230

DB 485 ArgValSerSerSerPheArgAspLysIleLeuGlnThrAsnPheSerProArgSer 504

QY 1231 ATTACAGATCATTTTCATGAGAAATTCAGAGTCCAGGACAGAGAAAGAAACAATGTGA 1290

DB 505 ThrIleAspSerPheSerAsnIleSerLysLys-----ArgAsnSerGluGluAlaAsn 522

QY 1291 GTCACACATCAAGAACAGAAAGAGAGATTCCAAATAATACATCCCACTAACTCCCTCCA 1350

DB 523 AspGluAsnAspGluThrAsnLeuLysIlePro-----IlePro 535

QY 1351 GAGAAAGTGGCTGGAACTCTTATTCAGGAAATAGAGAAATGCCTCGGTGGGATC 1410

DB 536 GluLysLysArgLysPheGlnGluValLeuGlnSerLysAsnIle----- 550

QY 1411 CATTTGCTCTCTTAAAGACCATCTCAAGTCAACAGTGCAGTGTTCGGGTAGACCTTCG 1470

DB 551 ---LeuValSerSerThrGluAspSerHisGluProValLysValThrGluAspSerGln 569

QY 1471 AGACGAGGAGAGGTTCCGAAACCTTTTCCAAACACATTAACAAAGATTTGTGGGATAACAAA 1530



570 ThrAlaIleHisValSerLys-----PheGluAspLeu---GluAsnLys 583  
 1531 CATGTGAATAAGCTTGTTCGGAACAAACTGTACCTGTG---GAAGATGAGATGGT 1587  
 584 SerMetGlu-----SerGluGlnSerLeuGlnLeuLeuSerGluSerGluAsnAsp 600  
 1588 GAGCACTACAGGAGTAGTGGAGCTCATTCAGACTGCACTTCTCAACAATTCACA 1647  
 601 AspLys-----ProLeuIleAspLeuLeuProLeuLeuAlaIleLys 614  
 1648 CGACCCAGCACTTGAAGAGTAGCTTCTGAATCAACAATGCGCATATCT----- 1698  
 615 ArgLysAsnLeuLeuValSerGlyValLeuGluLysGlyLysSerThrSerLys 634  
 1699 AAGAAATGGACCTTACAGCTTGTGTTGATTTCTGGATAGGACTTGAAGACGAG 1758  
 635 ThrLysPheAsp---ThrSerIleValAspPheIleGluLysProLysThrGlu----- 651  
 1759 GCCCAACATTTATCATCTGCTCATTTTACCTGACATGCTGAAATGCGACTGCTGCCA 1818  
 652 -----IleSerGluValLeuProGluGluLysArgLysAlaIleCys----- 665  
 1819 AATATTGTCACCCACCAATACCACTCTCTGAACAGAGATGAATCATTTCTGCACGATG 1878  
 666 -----AspGluSerGlnThrValArgValSerIleAspArgGlyValThrLys 681  
 1879 TCACAGGACAGATGCGGCTCTTTAGCTAATGCTTCTCTGCACATTTCCCGACGG 1938  
 682 ThrArgAsp---ValSerSerProValSerAspGlu-----LysSerGlu 695  
 1939 AATGCCAAGATGAATCGGATTTCTAGTTATACCCAGACATTAATCAATCGGTTGTT 1998  
 696 AsnValLeuHisGluGluAlaAsnSerGlyHisThrValMetAsnValHisSerSerLeu 715  
 1999 GAAGGACGTTTCATCAAGAAACCCAGAAACATGAAACACTCTTCTGCTACTTTCGAAGA 2058  
 716 AspProGlnProIleValGlnProAsnGluLeuGluSer---GlySerTyrLeuLysAsp 734  
 2059 GTCACAGAAAGAAA-----CCTACAGATGTTGTGACATTTACAGACAGCTTGAA 2112  
 735 LeuProAspArgAsnValGlyAsnSerGluLysValThrPheGlnGluAspAspIleAsn 754  
 2113 GATTTTCCAGATGGAAGAGGTGTGAAAGCTCTGACACAGCTTACACGCTCACTTACGAG 2172  
 755 Ser-----ProLysLeuGlnSerLysAsnAsnGlnThrValGlu 767  
 2173 GGTACCATAGAGGCAACGCGGCGCATGCTACAGGTGGATTTTGAAATCGTTTGT 2232  
 768 ---AlaValAsnThrGluThrSerAspLysLeuGlnGluLysGluAlaAsn----- 783  
 2233 CGAGTGTGTGCTGCTGCTGCGGACTTGTACAGAAGAAATCAGATTTTATCAATCCT 2292  
 784 -----HisGluLeuGluAsnIle 789  
 2293 GAATTCATGTTTTCAGGCTGTCTACTGAGTGG----- 2325  
 790 GluLysIleGluLysLeu---ThrGluValAspLysValSerLysSerAspAlaPhe 808  
 2326 -----CTGATCAATAGTGTCTTTATATATACAGGTACTGACAGTACAGT 2373  
 809 ProAspGlnGluIleLysAsnSerArgThrSerValGlnAsnGlyThrArgSerValSer 828  
 2374 GAATACACAGGCTATGCTCAACTTATCTCTGG-----CCGAAGCCATCAAGATGG 2427  
 829 LysAsnThrProGluLysGluThrLys---ValAspLysIleAspAsnValSerLysLys 848  
 2428 AGTGAAGACGACGATGGACGCGCGCTGCGAGATGCTGCGCATGACGCACTTCAC 2487  
 848 pValGluThrSerProGlySerCysGluThrSerSerAlaPheAlaLysThrTyrAlaG 868  
 2488 TTCAGACGCTACTCGATCATCTTGTGCTGTGAAAGTGAAGCTGAGCTTACAGGCT 2547  
 868 uLysGluValThrSerIleAsnLeu---Pro---SerValArgLysProLeuAspGluSer 886

2548 TAC-----TGCAGATTCTCTCCCTCCT 2568  
 887 TyrTyrAspHisSerIleSerProPheAspProLeuCysGlnSerSerPheLeuAlaPro 906  
 2569 GGAGTTCCTTCTGAAATCTTTCTGCAGTG 2598  
 907 GlnThrProValLysSerLysHisAlaLeu 916  
 RESULT 4  
 NP14 HUMAN  
 ID NP14 HUMAN STANDARD; PRT; 699 AA.  
 AC Q14978; Q15030;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa  
 DE nucleolar phosphoprotein) (Noppl40) (Nucleolar and coiled-body  
 DE phosphoprotein 1).  
 GN NOLC1 OR KIAA0035.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).  
 RC TISSUE=Leukemia;  
 RX MEDLINE=95386590; PubMed=7657714;  
 RA Pai C.-Y., Chen H.-K., Sheu H.-L., Yeh N.-H.;  
 RT "Cell-cycle-dependent alterations of a highly phosphorylated nucleolar  
 RT protein p130 are associated with nucleogenesis.";  
 RL J. Cell Sci. 108:1911-1920(1995).  
 RN [2]  
 RP SEQUENCE OF 3-699 FROM N.A. (ISOFORM BETA).  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=96051387; PubMed=7584026;  
 RA Nomura N., Miyajima N., Suzuki T., Tanaka A., Kawarabayashi Y.,  
 RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;  
 RT "Prediction of the coding sequences of unidentified human genes. I.  
 RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by  
 RT analysis of randomly sampled cDNA clones from human immature myeloid  
 RT cell line KG-1.";  
 RL DNA Res. 1:27-35(1994).  
 RN [3]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=96205319; PubMed=8630004;  
 RA Pai C.-Y., Yeh N.-H.;  
 RT "Cell proliferation-dependent expression of two isoforms of the  
 RT nucleolar phosphoprotein p130.";  
 RL Biochem. Biophys. Res. Commun. 221:581-587(1996).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=97168979; PubMed=9016786;  
 RA Chen H.-K., Yeh N.-H.;  
 RT "The nucleolar phosphoprotein p130 is a GTPase/ATPase with intrinsic  
 RT property to form large complexes triggered by P- and Mg2+.";  
 RL Biochem. Biophys. Res. Commun. 230:370-375(1997).  
 RN [5]  
 RP CHARACTERIZATION.  
 RX MEDLINE=20036810; PubMed=10567578;  
 RA Chen H.-K., Pai C.-Y., Huang J.-Y., Yeh N.-H.;  
 RT "Human Noppl40, which interacts with RNA polymerase I: implications  
 RT for RNA gene transcription and nucleolar structural organization.";  
 RL Mol. Cell. Biol. 19:8536-8546(1999).  
 CC -1- FUNCTION: Related to nucleogenesis, may play a role in the  
 CC maintenance of the fundamental structure of the fibrillar center  
 CC and dense fibrillar component in the nucleolus. It has intrinsic  
 CC GTPase and ATPase activities. May play an important role in  
 CC transcription catalyzed by RNA polymerase I.  
 CC -1- SUBUNIT: Interacts with RNA polymerase I 194 kDa subunit (RPA194)  
 CC and with casein kinase-II.  
 CC -1- SUBCELLULAR LOCATION: Shuttles between the nucleolus and the  
 CC cytoplasm. At telophase it begins to assemble into granular-like

pre-nucleolar bodies which are subsequently relocated to nucleoli at the early G1-phase.

-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=Alpha;  
IsoId=Q14978-1; Sequence=Displayed;  
Name=Beta;  
IsoId=Q14978-2; Sequence=VSP 004338;  
-!- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation cycles on CK2 and PKC sites. There is evidence suggesting that CD2 kinase phosphorylates p130 at the M-phase.

-!- SIMILARITY: Contains 1 Lish domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

EMBL; Z34289; CAA84063.1; -;  
EMBL; D21262; BAA04803.1; -;  
PIR; I38073; I38073.  
Genew; HGNC:15608; NOLC1.  
GK; Q14978; -;  
MIM; 602394; -;  
GO; GO:0005737; C:cytoplasm; TAS.  
GO; GO:0005730; C:nucleolus; TAS.  
GO; GO:0007049; P:cell cycle; TAS.  
GO; GO:0007067; P:mitosis; TAS.  
GO; GO:0006364; P:rRNA processing; TAS.  
InterPro; IPR006594; Lish.  
InterPro; IPR007718; SRP40\_C.  
Pfam; PF05022; SRP40\_C; 1.  
SMART; SM00667; Lish; 1.  
PROSITE; PS00896; Lish; 1.  
Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding;  
Alternative splicing; LISH.  
DOMAIN 10 42  
DOMAIN 84 566  
REPEAT 84 95  
REPEAT 125 136  
REPEAT 167 178  
REPEAT 221 232  
REPEAT 264 275  
REPEAT 325 336  
REPEAT 363 375  
REPEAT 425 436  
REPEAT 470 481  
REPEAT 519 529  
REPEAT 555 566  
DOMAIN 68 82  
DOMAIN 204 382  
DOMAIN 384 597  
DOMAIN 601 617  
MOD\_RES 563 563  
VARSEPLIC 241 241  
CONFLICT 3 3  
CONFLICT 133 133  
CONFLICT 291 292  
CONFLICT 456 456  
SEQUENCE 699 AA; 73720 MW; DFD4AD34EDF659FB CRC64;  
FTID=VSP 004338.  
D -> A (IN REF. 2).  
R -> S (IN REF. 2).  
YA -> SV (IN REF. 2).  
S -> P (IN REF. 2).  
K -> KWVITVSRAE (in isoform Beta).  
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
INTERACTS WITH RPA194.  
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
PHOSPHORYLATION (BY CK2) (BY SIMILARITY).  
K -> KWVITVSRAE (in isoform Beta).

Alignment Scores:  
Pred. No.: 0.000968  
Score: 166.50  
Percent Similarity: 34.58%  
Best Local Similarity: 22.10%  
Query Match: 2.46%  
DB: 1

US-09-302-812-5 (1-3814) x NP14\_HUMAN (1-699)

QY 47 GGGAGCCCTGCAGCA-----AAGCGCGCTGGGGCGCGCT-GGAACCT 87  
DQ 99 GlyProProAlaLysLysLysAlaAlaValProAlaLysArgValGlyLeuProGlyLys 118  
QY 88 TCTGCGCCGACTGCTCGGACTCCCGAGTTCCTGGCAGGCGAGAGCGCT----- 138  
DQ 119 AlaAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 136  
QY 139 -----GTTCTGACCCAGGAGCGCTCCGCTCAGTTCAGGCTCCCTCCGCTCCCTCCAGCC 195  
DQ 137 AspGluGluAspGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 156  
QY 196 TGCCTCTCGGCGCGGCGGAGCCGACAGAGGAGCCACCCACCTCGTTGTTTTC----- 249  
DQ 157 -----AlaLysAlaProProLysLysLysLysLysLysLysLysLysLysLys 173  
QY 250 -----AAACAAAGACTATTACTACTTGG 273  
DQ 174 SerSerGluAspGluProProLysAsnGlnLysProLysLysLysLysLysLysLys 193  
QY 274 ATGGATCTAAAGCA-----CCCAAGACAGCTGAATCAGAAAGTAAAGAAACCAATACA 330  
DQ 194 AlaGlnThrLysAlaProProLysProAlaArgAlaAlaProLysLysLysLysLys 213  
QY 331 AGAATTGACTCCATGATGAGTTCTGTGCAGAAAGATAACTTTTACCCACATAAGGTGAA 390  
DQ 214 AlaLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 233  
QY 391 AAATTGAAAATGTTCTCAGCTAAATCTTGATAAA----- 426  
DQ 234 LysAlaAlaAlaThrProLysLysLysLysLysLysLysLysLysLysLysLysLys 253  
QY 427 -----TCACCCACAGAAAGAGTTTACAGATTTTGAACCAACACAGACAG 471  
DQ 254 ValLysAlaAlaThrProThrArgLysSer-----SerSerGluAspSer 270  
QY 472 GCGAGTGTGCAAGTGGCAGATGAAGGAGATGACAGAACAGACTTTTGGCAAGTGA 531  
DQ 271 SerSer-----AspGluGluGluGlnLysLysLysLysLysLysLysLysLys 285  
QY 532 -----CCTCCCGCGGAGTCCGCTACCAAGACAGACTTAGTAA 570  
DQ 286 ProGlyProTyrSerTyrAlaProProSerAlaProProLysLys----- 302  
QY 571 GCTAACTTGTGCTCAGTCACCC----- 591  
DQ 303 ---SerLeuGlyThrGlnProProLysLysLysLysLysLysLysLysLysLysLys 321  
QY 592 -----CACACTGATGACACAGTGCACAGATCATGAAGAGACAGACAGACAG 645  
DQ 322 SerGluAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 338  
QY 646 TTTCTTACCTATTAATCTTGAATACAAAGCCACAGTAGGAGATGGGCGAGCCAGA 705  
DQ 339 -----ProProThrLysAlaValSerLysLysLysLysLysLysLysLysLys 356  
QY 706 ACCTACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGCTGTCAACAG 765  
DQ 357 Lys-----AlaAlaGluSerSerSerSerSerSerSerSerSerSerSerSer 370  
QY 766 GAGGAGGTGGATGTCTACACAGAGAGTCTTTTGTTCAGATGTTGGTGGCGAGACATGGA 825  
DQ 371 SerGluAspGluAlaProSerLysPro-----AlaGly 382  
QY 826 ACTGGACCAAAATGACAAACAAATTTGCTGACAAAGAAAGACAGCCTAGTGTTCGCT 885  
DQ 383 ThrThrLysAsnSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 398  
QY 886 CCATTGTGAGAAAGAGTGCAGTGCCTGAGTCACCAATG----- 921



QY 345 GATGAGTCTCTGTCAGAAAGATAAC-----TTTACCACATAAGGT 386  
 Db 263 uValAlaProAlaGlnProSerGlnThrPheProProSerLeuProSerLys 283  
 QY 387 GGAAAAATGGAATGTTCTCAGCTAAATCTT-----GATAATC 428  
 Db 283 rValAlaMetGlnGlnLysProThrAlaTyrValArgProMetAspGlyGlnAspGlnAl 303  
 QY 429 ACCACACAAAAG-----AGTTCACAGTATTGGAACCAACACACAGAC 470  
 Db 303 aProSerGluSerProGluLeuLysProLeuProGluAspTyr-----ArgGlnGlnTh 321  
 QY 471 T-----CCGAGTGTGTGCAAGTGCAGAAATGA 497  
 Db 321 rPheGluLysThrAspLeuLysValProAlaLysAlaLysLeuThrLysLeuLysMetPr 341  
 QY 498 AGGGAAGCATGACAGAAACAGCTTTTGGCAAGTGAAGTCCCGGGGAGCTCCGCTACCAA 557  
 Db 341 oSerGlnSerValGluGlnThrTyrSerAsnGluValHisCysValGluGluLeuLeu 361  
 QY 558 GCAGCTTAGTAACTGCTAATTTGGTCACTGACCCACACTGATGACCAAGTACAGACAGA 617  
 Db 361 sGluMetThrHisSer-----TtrProProLeuThrAlaLeuHisThrProSerTh 379  
 QY 618 TCATGAAGAAGACAGACACAATCAGCAGTCTTCTTACCTATATAAACTTGCAAAATACAA 677  
 Db 379 rAlaGluProSerLys-----PheProPheProThrLysAspSerGlnHisVa 395  
 QY 678 GCCAAGTAGAGATGGCCAGCCAGACCACTGAAGTCAAGTGGATCTCGCCAGTC 737  
 Db 395 lSerSerValThrGlnAsnGlnLysGlnTyrAsp-----ThrSe 408  
 QY 738 TGTGAAGACTGTACAGGTGTCTCAAGCAGGAGGTGGATGTCTACCAAGAGT---CC 794  
 Db 408 rSerLysThrHisSerAsnSerGlnGlnGlyThrSerSerMetLeuLysAspLeuGl 428  
 QY 795 TTTGTGATGTTGGTGGCCAGACATTTGGAATCGGACCAAAATATGACAAATTTGAC 854  
 Db 428 nLeuSerAsp-----SerGluAsp-----SerAspSerGluGlnTh 440  
 QY 855 TGCAAGAAAGACAGCCTAGTGTATTCGCTCAATTTGAGAAGAAAGTCAAGCTGAGTC 914  
 Db 440 rProGluLysProProSerSerSerAlaProProSerAlaProGlnSerLeuProGlu 459  
 QY 915 ACCAATGGATGTAGACAACTCGAAGAACAGTGTCTCAAGATTCAGACAGACAGAAAC 974  
 Db 460 -ProValAlaSerAlaHisSerSerSerAlaGluSerGluSerThrSerAspSerAspSe 479  
 QY 975 AAGTCCAGTCTTTGATGACAGACATGATCGTCTCTCCCAACAGCAAAATAAATCTTCAAG 1034  
 Db 479 rSerSerAspSerGluSerGluSerSerSerSerSerSerSerGluLysGlnLysProLeuGl 499  
 QY 1035 TTCCCAAGCAGAGACAGCTGATCGCATCTTAGAAACCG-----TATT 1079  
 Db 499 uThrProAlaProGluProGluProProThrThrAsnLysTtrpGlnLeuAspAsnTrpLe 519  
 QY 1080 GACTAAGGGAAGTGA-----GTAG 1100  
 Db 519 uThrLysValSerGlnProAlaAlaProProGluGlyProArgSerThrGluProProAr 539  
 QY 1101 ATTGCATTTTCCAAATTTGAAGAGAGAAATAATGCTGGACAGCTTAATATCCCAAG-- 1158  
 Db 539 gArgHisProGluSerLysGlySerSerAspSerAlaThrSerGlnGluHisSerGluSe 559  
 QY 1159 -----CCATCTGGAAACTCTTCTAGCCTT----- 1182  
 Db 559 rLysAspProProLysSerSerSerLysAlaProArgAlaProProGluAlaProHi 579  
 QY 1183 -----AATGTAGAGTGAAGTTC----- 1203  
 Db 579 sProGlyLysArgSerGlnLysSerProAlaGlnGlnGluProProGlnArgGlnTh 599

QY 1204 -----AAGCAG----- 1209  
 Db 599 rValGlyThrLysGlnProLysLysProValLysAlaSerAlaArgAlaGlySerArgTh 619  
 QY 1210 -----CATGGAAGAGGATTTCTAAATATACAGATCATTTTCATGAGATTTTCCAGTC 1262  
 Db 619 rSerLeuGlnGlyGluArgGluProGlyLeuLeuProTyrGlySerArgAspGlnThrSe 639  
 QY 1263 AGAGACACAGAAAGAAACAATGTGAAGTCAGACATCAAGAACACAGAAAGAAATTC 1322  
 Db 639 rLysAspLysProLysValLysThrLysGlyArgProArgAlaAlaLysSerAsnGluPr 659  
 QY 1323 AAAATACATCCCACTAACCTCCCTCCAGAGAAAGAGTGGTGGAACTCTTATTGAGCA 1382  
 Db 659 oLys-----ProAlaValProProSerSerGlu----- 668  
 QY 1383 AATGAGAAAAATCGCTCGGTGGATCATTTGCTTCTTAAAGACCATCTGCAAGTCA 1442  
 Db 669 -----LysLysLysHisLysSerSerLeuProAlaProSer----- 680  
 QY 1443 CACAGTGACTGTTCGGGTAGACCTTCTCAGACGACGAGGTTCCGAAACCTTTTCCAAC 1502  
 Db 681 -----LysAlaLeuSerGlyProGluProAlaLysAs 691  
 QY 1503 ACATTACAAAGATTTGTGGATAACAACATGTGAAATGCCTTCTCGAAACAAACTT 1562  
 Db 691 pAsnValGluAspArgThrProGluHisPheAlaLeuValProLeuThrGluSerGln-- 710  
 QY 1563 GTACCTCTGTGAAGATGAGAATGTGAGCGAACTCGAGAGTAGTGGAGGACTATTCA 1622  
 Db 711 -GlyProHisSerGlySerGlySerArgThrSerGlyCysArgGlnAlaValVal 730  
 QY 1623 GACTGCATCTTCAACAATTCACAGACCCACAGACTTGAAGGATCGATTCGAAATA 1682  
 Db 730 lGlnGluAspSerArgLysAspArgLeuProLeuProLeuArgAspThr----- 746  
 QY 1683 CAATGTGGCATATTCTAAGAAATGGACTTTACAGCTTTGGTGTGATTTCTGGGATAGST 1742  
 Db 747 -----LysLe 748  
 QY 1743 ACTTGAAGAAGCAGAGGCCCAACATTTATATCAGTCCATTTTACCTGACATGTTGAAAT 1802  
 Db 748 uLeuSerProLeuArgAspThrProProGlnSerLeu-----MetValLysIl 765  
 QY 1803 TCACCTCTGTCTGCCAAATATTTGCACCCAGCA----- 1836  
 Db 765 eThrLeuAspLeuLeuSerArgileProGlnProGlyLysGlySerArgGlnArgly 785  
 QY 1836 ----- 1836  
 Db 785 sAlaGluAspLysGlnProProAlaGlyLysLysHisSerSerGluLysArgSerSerAs 805  
 QY 1837 -----APACCACTCTGAAACAGAGATG-----NATCA 1865  
 Db 805 pSerSerSerLysLeuAlaLysArgLysGlyGluAlaGluArgAspCysAspAsnly 825  
 QY 1866 TTCTGTCAAGT-----TCACAGAACAGATCGCCAGTCTTTTAGCTAA 1910  
 Db 825 sLysileArgLeuGluLysGluLysSerGlnSerSerSerSerSerSerHisly 845  
 QY 1911 TGCTTTCTCTGCACATTTCCCGACGGAAATGCCAAGATCAATCGAGTATTTCTAGTTA 1970  
 Db 845 sGluSerSerLysThrLysProSerArgProSerSerGlnSerSerLysLysGluMetLe 865  
 QY 1971 CCCAGACATTAATCTCAATCGTGTGTGAGACCTTCAAGACCTTCAAGAAACCAAGAAAA-- 2028  
 Db 865 uProProProValSer-----SerSerSerGlnLysProAlaLysPr 880  
 QY 2029 -----CTGAAAAACACTCTTCTGCTACTTTTCAAGAGTCAACAG-----AAAAAAC 2075  
 Db 880 oAlaLeuLysArg-----SerArgArgGluAlaAspThrCysGlyGlnAspPr 896  
 QY 2076 TACAGGATTTGGTACATTTTACAGACAGAGTCTTGAAGAT-----TTTCCAGATGGGA 2129

```

Db      896 oProLysSerAlaSerSerThrLysSerAsnHisLysSerSerIleProLysGlnAr 916
      2130 AAGGTGTGAAAGCCTCTGACACGCTTACACGCTTACTTACAGAGGTACACATA----- 2181
      916 gArgValGluGlyLysGlySerArgSerSerGluHisLysGlySerSerGlyAspTh 936
      2182 -----GAGGCAACGCGCGA---GGCATGCTACA 2207
      936 rAlaLanProPheProValProSerLeuProAsnGlyAsnSerLysProGlyLysProGl 956
      2208 GGTGATTTTGCAAATCGTTTGTGGAGGTGCTGTGACTGTGCGGGACTT---GTACA 2264
      956 nValLysPheAspLysGln-----GlnAlaAspLeuHisMetAr 969
      2265 AGAAGAATCAGATTTTATCAATCCTCAATGATGTTTACGGCTG----- 2313
      969 gGluAlaLysLysMetLysGlnLysAlaGluLeuMetThrAspArgValGlyLysAlaPh 989
      2314 -----TTCACTGAGGTGCTGATCAATGATGCTTATTTATCAAGGTACTGAACA 2366
      989 eLysTyrLeuGluAlaValLeuSerPheIleGluCysGlyIleAlaThrGluSerGluSe 1009
      2367 GTACAGTGAA-----TACACAGGCTATGCTGAAACT-----TA 2399
      1009 rGlnSerLysSerAlaTyrSerValTyrSerGluThrValAspLeuIleLysPheI 1029
      2400 TCGTTGGGCGGAGCATGAGATGGAGTGGAGTGAAGAGGAGTGGCAG----- 2448
      1029 eMetSerLeuLysSerPheSerAspAlaThrAlaProThrGlnGluLysIlePheAlaVa 1049
      2449 -----CGCGCTGCGAGAGATCGTTGCTTGCATTTGACGCGATTCCTTACAGCGCTTACCT 2501
      1049 lLeuCysMetArgCysGlnSerIle-----LeuAsnMetAlaMetPheArgCysLysLy 1067
      2502 CGATCATGTTGTGCTGAGAAAGTACGCTGAGCTTAAACAGGCTTACTGGGATTCCT 2561
      1067 sAsp-----IleAlaIleLysTyrSerArgThrLeuAsnLysHisPhe----- 1081
      2562 CGCTCTGGAGTTCCTTCTGAAATCTTCTGAGTGGCCACGCGGAACTGGGGCTGTGG 2621
      1082 -----GluSerSerLysValAlaGlnAlaProSerProCysIl 1095
      2622 TGCCTTTGGGGT 2634
      1095 eAlaSerThrGly 1099

RESULT 6
ID      AKAG6 RAT
AC      Q9WVG7;
DC      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DE      A-kinase anchor protein 6 (Protein kinase A anchoring protein 6)
DE      (PKA6) (MAKAP).
DE      NCBI_TaxID=10116;
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-Sprague-Dawley;
RX      MEDLINE=9343692; PubMed=10413680;
RA      Kapiloff M.S., Shillace R.V., Westphal A.M., Scott J.D.;
RT      "MAKAP: an A-kinase anchoring protein targeted to the nuclear membrane
RT      of differentiated myocytes.";
RL      J. Cell Sci. 112:2725-2736(1999).
CC      -I- FUNCTION: Binds to type II regulatory subunits of protein kinase A
CC      and anchors/targets them to the nuclear membrane or sarcolemmal
CC      reticulum. May act as an adapter for assembling multiprotein
CC      complexes (By similarity).

```

```

CC      -I- SUBUNIT: Interacts with RII subunit of PKA, phosphatase 2B
CC      (calcineurin) and AKAP79 (By similarity).
CC      -I- SUBCELLULAR LOCATION: Sarcolemmal reticulum and nuclear membrane
CC      in heart muscle. Participation of multiple targeting signals allow
CC      correct intracellular targeting. These may be repeated motifs rich
CC      in basic and hydrophobic amino acids, palmitoylated/myristoylated
CC      motifs or alternatively splice targeting sequences.
CC      -I- DOMAIN: RII-alpha binding site, predicted to form an amphipathic
CC      helix, could participate in protein-protein interactions with a
CC      complementary surface on the R-subunit dimer.
CC      -I- SIMILARITY: Contains 2 spectrin repeats.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AF139518; AAD39150.1; -.
CC      InterPro: IPR002017; Spectrin.
CC      Pfam; PF00435; Spectrin; 3.
CC      SMART; SM00150; SPEC; 3.
CC
CC      KW REPEAT.
CC      FT REPEAT 768 847 SPECTRIN 1.
CC      FT REPEAT 1033 1148 SPECTRIN 2.
CC      FT DOMAIN 516 606 SER-RICH.
CC      FT DOMAIN 1558 1699 SER-RICH.
CC      FT DOMAIN 2082 2075 PKA-RII SUBUNIT BINDING DOMAIN.
CC      SQ SEQUENCE 2314 AA; 25439 MW; 2FB01DE0E2DDFA37 CRC64;
CC
Alignment Scores:
Pred. No.: 0.0018 Length: 2314
Score: 165.00 Matches: 145
Percent Similarity: 33.15% Conservative: 90
Best Local Similarity: 20.44% Mismatches: 265
Query Match: 2.44% Indels: 209
DB: 1 Gaps: 32
US-09-302-812-5 (1-3814) x AKAG6 RAT (1-2314)
QY 412 CTAAATCTTGTAATAATCAACCCACAGAAAGAGTTCACAGTATTGTAACCAACAGCAGACT 471
Db 1154 LeuAsnAlaAspHisGlnProMetGlnLeuIleValAsnLeuGluArgTrpGlu 1173
QY 472 GCGAGTGTG-----TGCAAGTGGCAGAAATGAA-----GGCAAGCAT 507
Db 1174 AlaIleValMetGlnAlaValGlnTrpGlnThrArgLeuGlnLysLysMetGlyLysGlu 1193
QY 508 GCAGAACAGCTTTTGGCAAGTGAAGTCCCGCGGGGACTCCGCTACCAAGCAGCTTAGT 567
Db 1194 SerGluThrLeuAsnValIleAspPro-----GlyLeuMet 1205
QY 568 AATGCTTAACATTGGTCAGTCACCCACACACTGATGACACAGTACACAGAT----- 618
Db 1206 AspLeuAsn---GlyMetSerGluAspAlaLeuGluTrpAspGluThrAspIleSerAsn 1224
QY 619 -----CATCAAGAA-----GACAGAGACAATCAGCAGTTCTTCT 651
Db 1225 LysLeuIleSerValHisGluGluSerAsnAspLeuAspGlnAspProGluProMetLeu 1244
QY 652 ACACCTATAAACTTGCAAAATATCAAAAGCCACAGTAGGAGATGGCGGAGCCAGCAACAC 711
Db 1245 ProAlaValLysLeuGluGluThr-----His 1253
QY 712 TGTAACTGCAGTGGATCTCGCCAGCTGTGAAAGACTGTACAGGCTGTCAACAGGAGGAG 771
Db 1254 HisLysAspSerGlyTyrGluGluGluAlaGlyAspCysGlySerProTyrThrSer 1273
QY 772 GTGGATGTCTACACAGAGAGTCTCT-----TTGCAGATGTGGT 810
Db 1274 AsnIleThrAlaProSerSerProHisIleTyrGlnValTyrSerLeuHisAsnValGlu 1293

```

QY 811 GCCGAGGACATTGGAAGTGGACCA-----AAATAATGACAAATAATTGACTGGCAAGAA 864  
 Db 1294 LeuHisGluAspSerHisThrProPheLeuLysSerProLysPheThrGlyThrThr 1313  
 QY 865 AGCAGCCTAGGTGATTCCTCCATTCCTGAGAAAGAGT----- 903  
 Db 1314 GlnProThrValLeuThrLysSerLysSerLysAspSerPheSerSerThrLysSer 1333  
 QY 904 -----GAGCCTGAGTCACCAATGATGATGATGAC 930  
 Db 1334 LeuProAspLeuLeuGlyGlySerGlyLeuValArgProTyrSerCysHisSerGlyAsp 1353  
 QY 931 AACTCGAAGAAACAGTTGTCAAGAT-----TCAGAGCAGATGAAAGAA----- 972  
 Db 1354 LeuSerGlnAsnSerGlySerGluSerGlyLeuValSerGluGlyAspAsnGluMetPro 1373  
 QY 973 -----ACAAGTCCAGTCTTTTCATGAGCAGAGATGATGCTTCCTCCCAACAGCAAT 1023  
 Db 1374 ThrAsnSerAspMetSerLysPheSerMetValAspGlySerProSerAsnProGluThr 1393  
 QY 1024 AAATCTTCAAGTTGCAAGCAAGAGAGTGTGCGATCTTAGGAAACGGTATTGACT 1083  
 Db 1394 GluHisProAspProGlnMetGlyAspAlaValLeuGluGlnLysPheLysAsp 1413  
 QY 1084 AAGGGAAGTGAAGTTAGATTCATTTCCATTTGCAATTTGAAGAGAGAAATPAATGCTGGGACCAAGT 1143  
 Db 1414 AsnGlyGluSerIleLysLeu-----SerSerValSer 1424  
 QY 1144 GACTTAATGCCAGCCA-----TCTGGAACACTCTTCTAGCCTTAAT----- 1185  
 Db 1425 ArgAlaSerValSerProValGlyCysValAsnGlyLysAlaGlyAspLeuAsnSerVal 1444  
 QY 1186 -----GTAGAGTGTAGAGTCTCCAGCAGCATGGAAGAGAGAT----- 1224  
 Db 1445 ThrLysHisThrAlaAspCysLeuGlyGluGluLeuGlnGlyLysHisAspValPheThr 1464  
 QY 1225 -----TCTAAAT-----ACAGATCATTTTCATGAGAATT 1254  
 Db 1465 PheTyrAspTyrSerTyrLeuGlnGlySerLysLeuLysLeuProMetIleMetLysGln 1484  
 QY 1255 TCCAGTCAGAGGACAGAGA-----AAA 1278  
 Db 1485 ProGlnSerGluLysAlaHisValGluAspProLeuLeuGlyGlyPheTyrPheAspLys 1504  
 QY 1279 GAAACATGTGAAGTCAGACATCAAGAGAGAGAGATTCCTCCAAATATACATCCACCT 1338  
 Db 1505 LysSerCysLysAlaLysHisGlnAlaSerGluSerGln-----Pro 1518  
 QY 1339 AACCTCCCTCCAGAGAGAGAGTGGCTGGGAATCCTTATGAGGAATAGAGAAATGCT 1398  
 Db 1519 AspAlaProProHisGluArgIleLeuAlaSerAlaProHisGluMetGlyArgSerAla 1538  
 QY 1399 CGG-----TGTGGGATCCATTTGCCCTTCCCTTAAGACCA 1431  
 Db 1539 TyrLysSerSerAspIleGluLysThrPheThrGlyIleGlnSerAlaArgGlnLeuSer 1558  
 QY 1432 TCTGCAAGTCACAGTCACTGCTTCGG----- 1458  
 Db 1559 LeuLeuSerArgSerSerSerValGluSerLeuSerProGlyGlyAspLeuPheGlyLeu 1578  
 QY 1459 -----GTAGACCTTCTGAGAGAGAGAGAGTTCGGAACCTTTTCCA 1500  
 Db 1579 GlyIlePheLysAsnGlySerAspSerLeuGlnArgSerThrSerLeuGluSerTrpLeu 1598  
 QY 1501 ACACATTACAAAGATTGTGGGATACAAACATGTGAAATGCTTGTTCGGAACAAAC 1560  
 Db 1599 ThrSerTyrLys-----SerAsnGluAsp 1606  
 QY 1561 TTGTACCTGTGGAAGATGAGAATGCT---GAGCGAAGTCAGGAGGTAGGTGGGAGCTC 1617  
 Db 1607 LeuPheSerCysHisSerSerGlyAspIleSerValSerSerGlySerValGlyGluLeu 1626

QY 1618 ATTGAGACTGCA-----CTTCTCAACAAATTACACGACCCCAAGAACTTGAAGGATCG 1671  
 Db 1627 SerLysArgThrLeuAspLeuLeuAsnArgLeuGlnSerProSerSerGlnGln 1646  
 QY 1672 ATTCTGAAATCAATG-----GCATATTCTTAAGAAATGGACTTT 1713  
 Db 1647 LysIleLysArgSerValSerAspMetThrLeuGlnSerSerGlnLysMetProPhe 1666  
 QY 1714 ACAGCT-----TTGGTTGATTTCTGGGATTAAGTACTTGAAGAGCAGAGCCCAACAT 1767  
 Db 1667 AlaGlyGlnMetSerLeuAspValAlaSerSerIleAsnGluAspSerProAlaSerLeu 1686  
 QY 1768 TTATATCAGTCCATTTTACCTGACATGGTGAATTTGCATCTCTCTGCCCAATATTGC 1827  
 Db 1687 ThrGluLeuSerSerSerAspGluLeu-----SerLeuCys 1698  
 QY 1828 ACCGAGCCAAATACCATCTCTGAAACAGAG----- 1857  
 Db 1699 SerGluAspIleValLeuHisLysAsnLysIleProGluSerAsnAlaSerPheArgLys 1718  
 QY 1858 ---ATGAATCATTTCTGCAGCATGTCACAGGACAGATGCGCAGCTCTTTAGCTAAAGCT 1914  
 Db 1719 ArgLeuAsnArgSerValAlaAspGluSerAspValAsnValSerMetIleValAsnVal 1738  
 QY 1915 TTCTTCTGCATTTCCCGACGAGNATCCCAAGATGAATCGAGTATTCTAGTTACCCA 1974  
 Db 1739 ---SerCysThrSerAlaCysThrAspAspGluAspAspSerAspLeuLeuSerSerSe 1757  
 QY 1975 GACATTAACTTCAATCGGT-----TGTTGAAGCAGCTTCATCAAGGAACACAGAA 2025  
 Db 1757 rThrLeuThrLeuThrGluGluGluLeuCysLeuLysAsp-----GluAspAspSe 1775  
 QY 2026 AAACGTGAAACACACTTCTGCTACTTTTCGAGAGATCGAG----- 2065  
 Db 1775 rSerIleAlaThrAspAspGluIleTyrGluGluSerAsnLeuMetSerGlyLeuAspTy 1795  
 QY 2066 ---AGAAAAAACCCTACAGGATTCG 2086  
 Db 1795 rIleLysAsnGluLeuGlnThrTrp 1803  
 RESULT 7  
 PCLO MOUSE  
 ID PCLO MOUSE STANDARD; PRT; 5038 AA.  
 AC Q9QX7; Q9QX6; Q9QZJ0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-  
 derived HLMN protein).  
 GN PCLO OR ACZ.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,  
 RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.  
 RC TISSUE=Brain;  
 RX MEDLINE=99439764; PubMed=10508862;  
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
 RA Kilimann M.W.;  
 RA "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
 zones, shares homology regions with rim and bassoon and binds  
 profilin.";  
 RL J. Cell Biol. 147:151-162(1999).  
 RN [2]  
 RP REVISIONS.  
 RA Kilimann M.W.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 4502-4682 FROM N.A.  
 RC TISSUE=Brain;  
 Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;





```

Db 1571 ArgPhelysThrIleGluLeuAsnSerThrValThrAspLysTyrSerAlaGluSerSer 1590
Qy 1093 GAAGTTAGA-----TTGCATTTCCAAATTGAAGAGAA----- 1125
Db 1591 GlnLysLysThrThrLeuTyrPheAspGluGluProGluLeuGluMetGluSerLeuThr 1610
Qy 1126 -----AATTAATCTGGACAGTGAATTAATGCTCAAG----- 1158
Db 1611 AspSerProGluAspArgSerArgGlyGluGlySerSerLeuHisAlaSerSerPhe 1630
Qy 1159 CCATCTGAAACTCT-----TCTAGCCTTAATGTAGAGTGTAGAGTTCCAAAG 1206
Db 1631 ThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSerAspSerPro 1650
Qy 1207 CAGCATGAAAGAGGATCTTAAA----- 1230
Db 1651 SerHisLysLysGlyGluSerLysGlnArgLysAlaArgHisArgSerHisGlyPro 1670
Qy 1231 -----ATTACAGAT----- 1239
Db 1671 LeuLeuProThrIleGluAspSerSerGluGluGluGluGluGluGluLeu 1690
Qy 1240 -----CATTTTCATGAGATTTTCCAAAGTCAGAG 1266
Db 1691 LeuLysGluGlnGluLysGlnArgGluLeuGluGlnArgLysSerSerSerLys 1710
Qy 1267 GACAGAAGAAAGAAATGTGAATGACATCAAGAAAGAAAGAAAGTTCAAAA 1326
Db 1711 LysSerLysLysAspLysAspGluLeuArgAlaGlnArg---ArgArgGluArgProLys 1729
Qy 1327 TACATCCACCTAAGCTCCCTCCAGAGAGAGTGGCTGGGAATCTTATGAGGAATG 1386
Db 1730 ThrProProSerAsnLeuSerProIleGluAsp-----AlaSerProThrGluGluLeu 1747
Qy 1387 AGAAATGCTCGG-----TGTTGGATTCATTTGCTTCC 1422
Db 1748 ArgGlnAlaAlaGluMetGluGluLeuHisArgSerSerCysSerGluTyrSerProSer 1767
Qy 1423 TTAAGACCATCTGCAAGTACACAGTACTGTTCGGGTAGACCTTCTGAGACGAGAG 1482
Db 1768 IleGluSerAsp----- 1771
Qy 1483 GTTTCGGAACCTTTTCCAAACATTAAGATTTGTGGAT---AACAAACATGTGAA 1539
Db 1772 ---ProGluGlyPheGluLeuSerProGluLysIleIleGluValGlnLysValTyrLys 1790
Qy 1540 ATGCTCTGTTTCGGAACAAACTGTGACCTGTGGAAGATGAGAATGGTGAGCAACTGCA 1599
Db 1791 LeuPro---ThrAlaValSerLeuTyrSerProThrAspGluGln----- 1804
Qy 1600 GGGAGTAGTGGAGCTCATTCAGACTGCACTTCAACAAATTCACAGACCCAGAAC 1659
Db 1805 -----SerValMetGlnLysGluGlyAlaGlnLysAla 1815
Qy 1660 TTGAAGATGCGATTCTGAAATACATATGCGCATATTTCTAAGAAATGGGACTTTACAGCT 1719
Db 1816 LeuLysSerAlaGluGluMetTyrGluGluMetMetHisLysProHisLysTyrLysAla 1835
Qy 1720 TTG-----GTTGATTTCTGGATAGGACTTGAAGAACGAGAG----- 1758
Db 1836 PheProAlaAlaAsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyMet 1855
Qy 1759 -----GCCCAACATTATATCATGCTCATTTTACCTGAC 1791
Db 1856 LeuIleGluAspTyrIleTyrGluSerLeuValGluAsp 1868

```

RESULT 8

PCLO\_HUMAN STANDARD; PRT; 5147 AA.  
ID PCLO\_HUMAN  
AC Q9Y6V0; O43373; O60305; Q9BYC8; Q9DIV2; Q9Y609;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)  
Piccolo protein (Acronin) (Fragments).  
PCLO OR ACZ OR KIA0559.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
SEQUENCE OF 1-759 FROM N.A.  
TISUE=Brain;  
RX MEDLINE=9433764; PubMed=10508862;  
RA Wang X., Kibschull M., Iane M.N., Lichte B., Petrasch-Parwez E.,  
RA Kilimann M.W.;  
RA "Acronin, a 550-kd putative scaffolding protein of presynaptic active  
RT zones, shares homology regions with rim and bassoon and binds  
RT profilin";  
RL J. Cell Biol. 147:151-162(1999).  
RN [2]  
SEQUENCE OF 552-4404 FROM N.A.  
RX Kraemer J., Wollam C., Wohldmann P., McGrane B.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).  
TISUE=Brain;  
RX MEDLINE=98290545; PubMed=9628581;  
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,  
RA Nomura N., Ohara O.;  
RA "Prediction of the coding sequences of unidentified human genes. IX.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro.";  
RL DNA Res. 5:31-39(1998).  
RN [4]  
SEQUENCE OF 4405-4439 FROM N.A.  
TISUE=Placenta;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Bhat N.K.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
SEQUENCE OF 4405-5147 FROM N.A.  
RX Kalicki J., Elliott G.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May act as a scaffolding protein involved in the  
CC organization of synaptic active zones and in synaptic vesicle  
CC trafficking (By similarity).  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By  
CC similarity).  
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
CC synaptic junctions (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=Q9Y6V0-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9Y6V0-2; Sequence=VSP\_003923, VSP\_003924, VSP\_003925,  
CC DT





QY	925	----	CTAGACAAC	TCGAGAAACAG	TTGTCAAGATT	TCAGAACGATG	AGAGAAACAAG	TC	979			
DB	1313	erLeuGlu	AlaGlnAlaSer	ThrLeuAlaAsp	GluLysSer	GluLysThr	GlnP	1333				
QY	980	CA-----	GTC	TTTGATGACAA	-----	GATGATCG	TTCTCC	CAACAGCAAAATAAA	1026			
DB	1333	roHisGlu	ValSerPro	GluGlnPro	LysAspGln	LysThr	GlnSerLeu	SerGluT	1353			
QY	1027	-----	CTT	CAAGTTGCCA	AGAGAACTG	ATGCGCAT	CTTACG	AAAC	1072			
DB	1353	hrLeuGlu	leThrIle	SerGluGlu	GluLeuLys	GluSerGln	GluGlu	---ArgLysA	1372			
QY	1073	GGTATT	TGCA	TAGGCAAG	TGAAGT	-----	-----	-----	1098			
DB	1372	spThrPhe	LysLysAsp	SerGlnAsp	ileProSer	LysAspHis	LysGluLys	1392				
QY	1099	-----	-----	-----	-----	AGATTG	CAATTTCC	AAAT	1114			
DB	1392	erGluPhe	ValAspAsp	IleThrThr	ArgGluPro	TyrAspSer	ValGluSer	1412				
QY	1115	TTCAAG	GAGAAATAAT	CTGGAC	CAGTGACT	TAAATGCC	AAGCCAT	CTGGAAC	CTCT	1174		
DB	1412	erGluSer	GluAsnSer	ProValPro	GlnArgLys	argThrSer	ValGlySer	SerS	1432			
QY	1175	CTAGCC	TTAATGTAG	AGTTTCC	AGCAGCAT	GGAAAA	AGGAT	TTCTAA	AATT	1233		
DB	1432	erSerAsp	GluTyrLys	GlnGluAsp	SerGlnGly	SerGlyGlu	GluGluAsp	PheIleA	1452			
QY	1234	-----	-----	-----	-----	ACAGAT	CAAT	1243				
DB	1452	rgLysGln	IleGluMet	SerAlaAsp	GluAsp	AlaSerGly	SerGluAsp	AspGluP	1472			
QY	1244	TCATG	GAGA-----	ATTTCC	AAATCAG	TCAGAGCA	GAGACAG	AGAAAGACA	CAAT	1285		
DB	1472	hele	argAsnGln	LeuLysGlu	LysSerSer	ThrGluSer	GlnLysGlu	LysGluT	1492			
QY	1286	GTGAA	-----	GTC	CAGCATCA	AAAGAAC	AGAAAG	AGATTC	CAAA	1324		
DB	1492	hrLysGly	LysGlyLysIle	ThrAla	GlyLeHis	argArgLeu	ThrArgLys	SerSerT	1512			
QY	1325	AATCAT	TCCACCT	CACTCC	TCAGAG	AGAAAGT	GGTGGCA	ATCTCT	TATTGAG	CAAAA	1384	
DB	1512	hrSerIle	AspGluAsp	AlaGlyArg	argHis	serTrpHis	AspGluAsp	AspGlu	AlaP	1532		
QY	1385	TGAGAAA	ATATGC	CTCGGTG	GGGATCC	ATTGCT	CTCTTA	AGACCAT	TCGCAAG	TCACA	1444	
DB	1532	heAspGlu	SerProGlu	LeuLysTyr	ArgGluThr	LysSerGln	GluSerGlu	LysGlu	LeuV	1552		
QY	1445	CAGTGA	CTGTTCCG	GTAGACT	CTCTG	AGAA-----	GCAGAG	AGGTTCC	GAAAC	CTTTTC	1498	
DB	1552	alValThr	GlyGlyGlyLeu	argargPhe	LysThrIle	GluLeu	AsnSerThr	IleA	1572			
QY	1499	CAACAC	ATTAACA	GATTGTG	GGATAAC	AAACATGT	GAAATGC	CTTGTTC	GGGAACA	1557		
DB	1572	laAspLys	TyrSerAla	GluSerSer	GlnLysLys	ThrSerLeu	TyrPhe	AspGlu	GluP	1592		
QY	1558	-----	AACT	TGACCTGT	GGAAGAT	GAGAAATG	TGACG	CAACAT	CTGCGG	AGTAGGT	1609	
DB	1592	roGluLeu	GluMetGlu	SerLeuThr	AspSerPro	GluAsp	ArgSerArg	GlyGlu	Lys	1612		
QY	1610	GGGAG	CTCATTC	CAGACTG	CACTTCT	CAACAA	ATTTCA	CAGAC	CCCAACT	GTG	1662	
DB	1612	erSerSer	LeuHisAla	SerSerPhe	ThrProGly	ThrSerPro	GlyThrSer	ThrSer	ValSer	SerL	1632	
QY	1663	-----	AAAG	ATCGCAT	CTGTA	AAATACA	ATGTGG	CATATTT	CTAAG	AAATGG	GACT	1711
DB	1632	euAspGlu	AspSerAsp	SerSerPro	SerHis	LysLysGly	LysSerLys	-----	-----	-----	1648	
QY	1712	TTACAG	CTTTGG	TGATTT	CTCGG	ATAGG	TACTT	GAGAC	AGACAG	CGCCCA	CAATTT	1771
DB	1649	-----	-----	-----	-----	GlnGln	ArgLysAla	Arg	1657			

1772 ATCAGTCCATTTTACCTGCATCGTGGTGAATAATTGCACCTCTGTCTGCGCAATATT----- 1822  
1657 roHisGlyProLeu-----LeuProThrIleGluAspS 1668  
1825 -----TGCACCCAGCCCAATACCCTCTCTGAAA---CAGAAGATGAATC 1864  
1668 erSerGluGluGluLeuArgGluGluGluGluLeuLeuLeuGluGluGluGluGlnA 1688  
1865 ATTCTGTCAGCATGTACAGGAACACAGATGCCAGTCTTTTAGCTAATAGCTTTCTCTGCA 1924  
1688 rgGluIleGluGluGlnArgLysSerSerSer----- 1699  
1925 CATTTCCCGACGGAATGCCAAGATGAATTCGGAGTATTCTAGTACCAGACATTAACT 1984  
1700 -----LysLysSerLysLysAspLysAspGluLeu----- 1709  
1985 TCAATCGTGTGTTGAAGGAGCTTCATCAAGGAAACACAGAAAACCTGAAAACACTCTTCT 2044  
1710 -----ArgAlaGlnArgArgArgGluArgProLysThrProProS 1723  
2045 GCTACTTTTGAAGAGTACAGAAAAAACTCAGAGATTGGTGACATTTACAAGACAGA 2104  
1723 erAenLeuSerProIleGluAspAlaSerProThrGluGluLeu-----ArgGlnA 1740  
2105 GTCCTTGAAGATTTTCCAGATGGGAAGGTGTGAAAAGCTCTGCACGCTTACACGTCA 2164  
1740 laAlaGlu-----MetGluGluLeuHisArgS 1749  
2165 CT-----TACAGGGTACCATAGACGCAACGCCGAGGCGATCTACAGGTGG 2212  
1749 erSerCysSerGluTySerProSerIleGluSerAspProGluGly----- 1764  
2213 ATTTTGCAAATCGTTTGTTCGAGTGGTGTGACTGTGTGCGGAGCTGTGTACAAGAGAAA 2272  
1764 ----- 1764  
2273 TCAGATTTTAAATCAATCCTGAATGATGTTTCA-----CGCTGTTCA 2317  
1765 -----PheGluIleSerProGluLysIleIleGluValGlnLysValTyLysLeuProT 1783  
2318 CTGAGGTG-----CTGGATCACAAATG 2338  
1783 hrAlaValSerLeuTySerProThrAspGluGlnSerIleMetGlnLysGluGlySerG 1803  
2339 AGTGTCTATTATCACAGGTACTGAAACAGTACAGTGAATACACAGGCTATGTCTGAAACTT 2398  
1803 InLysAlaLeuLysSerAlaGluGluMetTyTyGluGluMetMetHisLysThrHisLysT 1823  
2399 ATCGTGGGCCCGAAGCCATGAAGATGGGAGTGAAGAAGGAGGATTCGCACGGCGGCTGCA 2458  
1823 yr-----LysAlaPheProAlaAlaAenGluArgAsp----- 1833  
2459 CGGAGATCGTGGCCATTACAGCACTTCATCTCAGACGCTACTCGATCAGTTTGTGCTG 2518  
1834 --GluValPheGluGlnProLeuTyTyGlyMetLeuIleGluAspTyTyIleTyG 1853  
2519 AGAAAGTCAGACGTCAGCTTAAACAAGGCTTACTCGGATTCCTCCGTCCTGGAGTTCCTT 2578  
1853 luSerLeu-----ValGluAspThrTyArgGly-----SerValAspG 1866  
2579 CTGAAATCTTTCTGCAGTGGCCACGGGAAACTCGGGCTGTGGTCCCTTGGGGGTCAAG 2638  
1866 lySerLeuLeuThrArgGlnGluGluAenGlyPheMetGlnGlnLysGlyArgGluG 1886  
2639 CTAGATTTAAAGCCTTAATACAGATC----- 2664  
1886 InLysIleArgLysSerGluGlnIleTyGluAspProMetGlnLysIleThrAspLeuG 1906  
2665 -----CTGGCAGCTGTGCGGCTGAACCTGACCTGCTT 2698  
1906 InLysGluPheTyTyGluLeuGluSerLeuHisSerValProGluGluAspIleValS 1926  
2699 ATTTACCTCTTTGGGACCTCAGAGTTGTATGAGACACATTTACAGCATGCACATTTCTT 2758

```

Db      1926  exSerSerPheillelleProGlu---SerHisGluilleValaAspLeuGlyThrMetValt 1945
QY      2759  CC-----GAGAGGAG---CTGATCTTGGAAAAGGTACAAAGTTATTCTTAA 2803
Db      1945  hrSerThrGluGluGluArgLysLeuLeuAspAlaAspAlaAlaTyrgluGluLeuMetL 1965
QY      2804  GATACATACATGAGAAATGACAGAAATGTTCCACCCCTGGACCCAGCATCAAGCTTTATC 2863
Db      1965  ysArgGlnGlnMetGlnLeu-----ThrProGlySerSerProThrGlnAlaP 1981
QY      2864  CATTATATACCATCTGTTGACTCAAGTCAGAGACCACTGACATGCCA 2913
Db      1981  roileGlyGluAspMetThrGluSerThrMetAspPheAspArgMetPro 1997

RESULT 9
NKCR_HUMAN
ID NKCR_HUMAN STANDARD; PRT; 1462 AA.
AC P30414;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE related protein) (NK-TR protein).
DE NKTR.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93133824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA Ortaldo J.R.
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells."
RL Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
[2]
RP REVISIONS.
RA Anderson S.K.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Component of a putative tumor-recognition complex.
CC Involved in the function of NK cells.
CC -!- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC via its N-terminus.
CC -!- SIMILARITY: Contains 1 cyclophilin-like ppiase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04288; AAA35734.2; -.
CC EMBL; AF184110; AAD56402.1; -.
CC PIR; A47328; A47328.
CC HSSP; Q27450; 1A33.
CC Genew; HGNC:7833; NKTR.
CC MIM; 161565; -.
CC GO; GO:0016018; F:cyclosporin A binding; TAS.
CC InterPro; IPR002130; CSA_Ppiase.
CC Pfam; PF00160; pro isomerase; 1.
CC PRINTS; PR00153; CSAPPISMRASE.
CC PROSITE; PS00170; CSA_PPIASE_1; 1.
CC PROSITE; PS00072; CSA_PPIASE_2; 1.
CC Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
CC DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
CC FT FT 219 240 ARG/LYS-RICH (BASIC).
CC FT FT 421 457 ARG/LYS-RICH (BASIC).
CC FT FT 970 1010 ARG/LYS-RICH (BASIC).

```

```

FT DOMAIN 194 244 ARG/SER-RICH.
FT DOMAIN 466 574 ARG/SER-RICH.
FT DOMAIN 664 814 ARG/SER-RICH.
FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.
SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;

Alignment Scores:
Pred. No.: 0.0032 Length: 1462
Score: 160.50 Matches: 143
Percent Similarity: 33.47% Conservative: 102
Best Local Similarity: 19.54% Mismatches: 281
Query Match: 2.37% Indels: 206
DB: 1 Gaps: 30

US-09-302-812-5 (1-3814) x NKCR_HUMAN (1-1462)
QY 85 ACTTGTGCGCCGACCTCCCTGGACTCCCGAGCTCCCTGGCAGGCGAGGCGGTGTTCTC 144
Db 537 ThrAlaSerLysSerSerHisSerArgSer----- 547
QY 145 GACCCCAAGACGCTCCCGTCAGTTCCAGGTCCCTCCGTCCTCGCAGCTCGCTCTCG 204
Db 548 -----ArgSerLysSerArgSerSer 554
QY 205 GCGCGCGCGGCGCCGACAGAGGCAACGCCACCTCGTTGTTTCAAAACAAAAGACTATT 264
Db 555 SerLysSerGly---HisArgLysArgAla----- 563
QY 265 ACTACTTGGATGATATAAGGACCCAGACAGCTGAATCAGAAAGTAAGAAACAAC 324
Db 564 -----SerLysSerProArgLysThrAlaSerGlnLeuSerGluAsnLys 578
QY 325 AATCAGAAATTCACCTCCATGATGAGTCTGTG---CAGAAAGATTAATTTACCCACAT 381
Db 579 ProValLysThrGluProLeuArgAlaThrMetAlaGlnAsnGluAsnValValGln 598
QY 382 AAGGTGMAAAATGGAATAATGTTCTCAGCTAAATCTTGATAAATCACC----- 432
Db 599 ProVal---ValAlaGluAsnLysProValLysProLeuSerAspSerProProSer 617
QY 433 -----ACA 435
Db 618 ArgTrpLysProGlyGlnLysProTrpLysProSerTyrgluArgIleGlnGluMetLys 637
QY 436 GAAAGAGTTTACAGATATTTGAACCAACAGCAGCTGCGAGTGTGTGCAAGTGCGCAAT 495
Db 638 AlaLysThrThrHisLeuLeuProLysGlnSerThrTyrsSerLeuAlaAsnLysGlu 657
QY 496 GAAAGGGAAGCATGCAGAACAGCTTTTGGCAAGTGAGCCTCCCGGGGAGCTCCGTACCA 555
Db 658 ThrGly-----SerSerSerSerTyHis 665
QY 556 AAGCAGCTAGTATGCTTAACATTTGTCAGTCACTCCACACACT-----GATGACCAAGT 609
Db 666 LysArgGlyLysAsnSerGluSerAspGlnSerThrTyrsSerLysTyrsSerAspArgSer 685
QY 610 GACACAGATCATCAAGAGACAGAGAC-----AATCAGCAG 645
Db 686 SerGluSerSerProArgSerArgSerArgSerArgSerArgSerArgSerArgSer 705
QY 646 TTTCTTACACCTATAAACTTGAATGCAAAATACAAAGCAACAGTAGAGATGGCGAGCCAGA 705
Db 706 TyrThrArgSerArgSerLeuAlaSerSer-----HisSerArgSerArg 720
QY 706 AGCACTGTAGTGCATGATGATTCGCCACTCTGTGAAGACACTGTACAGGCTGTCAACAG 765
Db 721 SerProSerSerArgSerHisSerArgAsnLysTyrsSerAspHisSerGlnCysSerArg 740
QY 766 GAGGAGGTGATGTGCTACCAAGAGAGTCTTTTGCAGATGTTGTGTGCGGAGGACATTGGA 825
Db 741 -----SerSerSerTyThrSerLysSerSerAspAsp---Gly 752
QY 826 ACTGGACCAAAAATATGACAAATATGCTGGACAAAGAACGCCCTAGGTGATTCCGCT 885

```



DB: 1 13  
Gaps:

US-09-302-812-5 (1-3814) x DSPP\_RAT (1-687)

QY	73	GCGCGCTGGAGACTTCTGC	CGCGCAGTCCGCGACTCCG	GAGCTTCCTCGCAGGCAG	132
Db	309	GlyGlyAspAsnThrSerSer	GlulThrGlnGlyIleGluGly	AspGlyThrGln	328
QY	133	AGGCGTGTTCTCGACCCC	AAGACGCTCCGTC	CAGTTCAAGGTCCCTCGTC	CGCCA 192
Db	329	ThrThrGlnAspAsnGln	AsnLeuSerProThrGluGly	GlyIleSerGlnAlaGlu	348
QY	193	GCGTGGTTCGCGGCGGCG	-----	GGACCGCACAGAGC	228
Db	349	AlaCysProSerGlyGlnSer	GlnAsnGlnGlyLeuGluThr	GlnGlySerSerThrGly	368
QY	229	AAGCCACTCGTTTGT	TTCATAAACAAGAATATTACT	TGTGATGATCATTAAGA	288
Db	369	AsnLysSerSerIleThrLys	GlusGlyLeuSerGlySerLys	AspSerAsnGly 388	
QY	289	-----	CCCAAGACAGCTGAATC	AGAA-----	309
Db	389	HisHisGlyMetGluLeu	AspLysArgAsnSerProLys	GlnGlyLeuSerAspLysPro	408
QY	310	-----	AGTAAGAACAACAAT-	-----	342
Db	409	GlnGlyAlaGluLysSer	AspThrHisAsnAsnMetGly	HisSerArgIleGlySer	428
QY	343	ATGATCAGTTCTGTG	CAGAAAATACATTTTACC	ACATTAAGGTGGAAAAAT	TGGAAAT 402
Db	429	SerSerAsnSerAspGly	HisAspSer--TyrAspPhe	AspAspGluSerMetGlnGly	447
QY	403	GTTCCTCAGCTAATCT	TGATAAATCACCACAGAAA	GAGTTCACAGTATTGGA	ACCAA 462
Db	448	AspAspProAsnSerSer	AspGluSerAsnGlySer	AspGlySerAspAlaAsnSer	467
QY	463	CACGAGACTCGAGTGT	GTCAGTGCAGATGAAGG	GAAGCATCAGAACACGCTTT	TG 522
Db	468	GluSerAlaIle-----	-----	GluAsnGlyAsnHisGly	AspAlaSerTyr 481
QY	523	GCAAGTGAGCCTCCCG	GGGAGCTCCGCTACCA	AGACGCTTAGTAATGCTTA	CATTGGT 582
Db	482	ThrSerAspGluSerSer	AspAsnGlySerAsp-	-----	SerAspSerHisAlaGly 498
QY	583	CAGTCCCCACACTGAT	GACACACAGTGCACAGAT	CATGAAGAAGACAGACA	CAATCAG 642
Db	499	GluAspAspSerSerAsp	AspThrSerAspThrAsp	AspSerAspSerAsnGly	AspAsp 518
QY	643	CAGTTCTTACACCTA	TAATAAACTTGCAAA	TACAAGCCACAGTAGAG	ATGGCAGGCC 702
Db	519	AspSerGluSerLys	AspLys-----	-----	AspGluSerAsp 529
QY	703	AGAAGCAACTGTGA	GTGATCTCGGCAGTCT	CTGTGNAAGACTGTAC	AGCTGTCAA 762
Db	530	AsnSerAsnHisAsp	AsnAspSerAspSerGlu	SerLysSerAspSerSer	AspSerAsp 549
QY	763	CAGCAGGAGGTGGAT	GTGTACCAGAGTCTTT	TGTTCATGTTGTCGCG	ACGACATT 822
Db	550	SerAspSerSerAsp	SerAspSerAspSerSer	AspSerAsp--SerSer	GluThrSer 568
QY	823	GGAACTGGACCA	MAAATATGACAAATTC	AGTCTGGACAGAACG	ACCTAGTGTTCG 882
Db	569	AspSerSerAspSer	AspThrSerAspSerSer	AspSerSerSerSerSer	AspSerSer 588
QY	883	CTCCATTGGAAGA	AGAAAGTGACCTGAT	CACCAATGATGTAG	ACAACCTCGAGAAAC 942
Db	589	SerAsnSerSerAsp	ThrSerAspSerSerAsp	SerSerSerSerSerSer	AspGlyAspSerAspGly 608
QY	943	AGTTGTCAGATTC	ACAAGCAGATGAGAA	CAACAGTCCAGTCTT	TGTGATGACGACAGATGAT 1002
Db	609	AspSerSerAspSer	AspSerSerAspSerSer	AspSer-----	SerAsnSerSerAsp 625
QY	1003	CGTTCTCCCCAA	CAGCAATAAACTTTC	AGTTGTCGCAAG	CAAGCTGATGGCGAT 1062

448	AspAspProAsnSerSerAspGluSerAsnGlySerAspGlySerAspAspAlaSer	467
463	CAGCAGACTCGAGTGTGTGCAAGTGGCAGATGAGGGAAGCATCGACAGAACAGCTTTTG	522
468	GluSerAlaIle-----GluAsnGlyAsnHisGlyAspAlaSerTyr	481
523	GCAAGTCAGCTCCCGGGGACTCCGCTACCAAGCAGCGTTAGTAATGCTAACATGGT	582
482	ThrSerAspGluSerSerAspAsnGlySerAsp-----SerAspSerHisAlaGly	498
583	CAGTCACCCCACTGATGATGACCACTGACACACAGATCATCAAGAAGCAGACAGACAATCAG	642
499	GluAspAspSerSerAspAspThrSerAspAspSerAspSerAsnGlyAspAsp	518
643	CAGTTCTTACACCTATAAACTTGCAAATACAAAGCCACAGTAGGAGATGGCGAGGCC	702
519	AspSerGluSerIysAspIys-----AspGluSerAsp	529
703	AGAAGCAACTGTTAAGTGCATGGATCTCGGCATCTGTGAAACACTGTACAGCTGTCAA	762
530	AsnSerAsnHisAspAsnAspSerAspSerGluSerIysSerAspSerSerAspSerAsp	549
763	CAGGAGGAGGTGGATGTCCTACCAGACAGTCCTTTGCATGCTGTGTCGCCAGGACATT	822
550	SerAspSerSerAspSerAspSerAspSerAspSerAspSerAspSerAspSerAsp	568
823	GGAACTCGACCCAAAATAATGACAAATTCACCTGGACGACAGAAAGACGACCTAGTGATTG	882
569	AspSerSerAspSerAspThrSerAspSerSerAspSerSerAspSerSerAspSer	588
883	CCTCCATTGAAAGAAAGTAGCAGCTGAGTCACCAATGCTATGACCAACTCGAGAAAC	942
589	SerAsnSerSerAspThrSerAspSerSerAspSerSerAspGlyAspSerSerAspGly	608
943	AGTTGTCAAGATTACGAAGCAGATGAGAAACAAAGTCCAGTCTTTGATGACGAAGATGAT	1002
609	AspSerSerAspSerAspSerSerAspSerAspSerAspSer-----SerAsnSerSerAsp	625
1003	CGTTCTCTCCAAACAGCAATAAATCTTTCAGTTGGCCAAAGCAGAGAAAGCTGATGGCGAT	1062







01-FEB-1994 (Rel. 28, Last sequence update)  
 10-OCT-2003 (Rel. 42, Last annotation update)  
 ENL protein.  
 MLLT1 OR ENL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93046667; PubMed=1423624;  
 RA Tkachuk D.C., Kohler S., Cleary M.L.;  
 RT "Involvement of a homolog of Drosophila trithorax by 11q23  
 chromosomal translocations in acute leukemias";  
 RL Cell 71:691-700(1992).  
 RP [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=94362221; PubMed=8080983;  
 RA Rubnitz J.E., Morrissey J.J., Savage P.A., Cleary M.L.;  
 RT "ENL, the gene fused with HRX in t(11;19) leukemias, encodes a  
 nuclear protein with transcriptional activation potential in lymphoid  
 and myeloid cells.";  
 RL Blood 84:1747-1752(1994).  
 CC -!- FUNCTION: Capable of activating transcription from synthetic  
 reporter genes in both lymphoid and myeloid cells.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- DISEASE: Involved in acute leukemias through a chromosomal  
 translocation t(11;19)(q23;p13.3) that involves MLL/HRX and MLLT1.  
 CC The result is a rogue activator protein.  
 CC -!- SIMILARITY: TO HUMAN MLLT3/AF9.  
 CC -!- SIMILARITY: TO YEAST TRANSCRIPTION INITIATION FACTOR TFIIF SMALL  
 SUBUNIT (TFG3/ANCL).  
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; ENL.html".  
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/ENL.html".  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L04285; AA58457.1; --  
 CC FIR; B44265; B44265.  
 CC Genew; HGNC:1134; MLLT1.  
 CC MIM; 159556; --  
 CC GO; GO:0005634; C:nucleus; TAS.  
 CC GO; GO:0003677; F:DNA binding; TAS.  
 CC GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.  
 CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
 CC InterPro; IPR005033; YEARS.  
 CC Pfam; PF03366; YEARS; 1.  
 CC Transcription regulation; Activator; Nuclear protein;  
 CC Chromosomal translocation; Proto-oncogene.  
 CC FT DOMAIN 268 280 POLY-PRO.  
 CC FT DOMAIN 319 325 POLY-SER.  
 CC FT DOMAIN 374 389 POLY-SER.  
 CC FT DOMAIN 457 460 POLY-PRO.  
 CC FT DOMAIN 463 466 POLY-PRO.  
 CC SEQUENCE 559 AA; 62082 MW; A6836722EC84D004 CRC64;  
 Alignment Scores:  
 Pred. NO.: 0.00381 Length: 559  
 Score: 157.50 Matches: 110  
 Percent Similarity: 34.60% Conservative: 63  
 Best Local Similarity: 22.00% Mismatches: 205  
 Query Match: 2.33% Indels: 122  
 DB: 1 Gaps: 20

US-09-302-812-5 (1-3814) x ENL\_HUMAN (1-559)

52 CCCTGCGACGAAGCGCGTGG-----GGCGCGCTGGAACTTTTCGG---CCGACT 99

Db 127 ProThrThrGluPheArgTyrLysLeuLeuArgAlaGlyGlyValMetValMetProGlu 146  
 QY 100 GCCTCGGACTCCCGAGCTTCCTCGGAGGAGGCGGTCTCTCGACCCCAAGAACGCT 159  
 Db 147 GlyAlaPheThrValSerArgProSer-----ProAspTyr 158  
 QY 160 CCCTCCAGTTTCAGGTCCTCCCTCGCTCGGAGGCTCGCTCGGCGGGCGGACCG 219  
 Db 159 ProMetLeuProThrIleProLeuSer----- 167  
 QY 220 CACAGAGCAAGCCACCTCCCTCTTTTCAACAAAGAGACTATTACTACTTGGATGAT 279  
 Db 168 -----AlaPheSerAspProLysLysThr-----LysPro 177  
 QY 280 ACTAAGAGCAAGCCACAGAGCTGAATCAGAAAGTAAAGAAACAAACATACAAAGATTGAC 339  
 Db 178 SerHisGlySerLysAspAlaAsnLysGluSerSerLysThrSerLys----- 193  
 QY 340 TCCATGATGAGTTCTGTGCAGAAAGATACTTTTACCCACATAGCTGGAAGAAATGGAA 399  
 Db 194 -----ProHisLysValThrLys-----Glu 200  
 QY 400 AATGTTCTCAGCTAAATCTTGATAAATCACCACAGAAAGAGTTCACAGTATTTTGAAC 459  
 Db 201 HisArgGluArgProArgLysAsp---SerGluSerLysSerSerSerLysGluLeuGlu 219  
 QY 460 CAACAGCAGACTCGCAGT-----GTGTCAAAGTGCAGAGTAAAGGAAGCAATGCA 510  
 Db 220 ArgGluGlnAlaLysSerSerLysAspThrSerArgLysLeuGlyGluArgLeuPro 239  
 QY 511 GAACAGCTTTTGGCAAGTGAAGCTCCCGCGGAGTCCCGCTACCAAG----- 558  
 Db 240 LysGluGluLysAlaProProLysAlaAlaPheLysGluProLysMetAlaLeuLys 259  
 QY 559 -----CAGCTTAGTAATGCTACATTTGTCAGTCCAGTCCACCCACACTGATGACCAC 606  
 Db 260 GluThrLysLeuGluSerThrSerProAsnProGlyProProProProProPro 277  
 QY 607 AGTGACACAGATCATGACAGACAGACAGACAATCAGCAGTTTCTTACACTATAAACTT 666  
 Db 278 -----ProProProArgAla 282  
 QY 667 GCAATACAAAGCCACAGTAGGAGATGGG-----CAGGCAGCA 705  
 Db 283 SerSerLysArgProAlaThrAlaAspSerProLysProSerAlaLysLysGlnLysLys 302  
 QY 706 AGCAACTGTAAGTGCAGTGGATCTCGCCAGTCTGTGAAAGACTGTACAGCTGTCAACAG 765  
 Db 303 SerSerSerLysGlySerArgSerAlaProGlyThrSerProArgThrSerSerSer 322  
 QY 766 GAGGAGGTGATGTGCTACAGAGAGTCTTTGTGAGATGTTGGTCCGAGGACATT--- 822  
 Db 323 SerPheSerAspLysLysProAlaLysAspLysSerSerThrArgLysGluLysValLys 342  
 QY 823 ---GGAACTGGACCAAAAATGACAA---TTGACTGCAACAAGAACAGCAGCTAGGT 876  
 Db 343 AlaGluSerGluProArgGluAlaLysLysAlaLeuGluValGluLysSerAsnSerGlu 362  
 QY 877 GATTTCGCTCCATTGAGAAAGAAAGTGAAGTCCAGTACCAATGGATGTAGACAACTCG 936  
 Db 363 AspGluAlaSerPheLysSerGluSerAlaGlnSerSerProSerAsnSerSerSer 382  
 QY 937 AGAAACAGTTGTCAAGATTGAGAA-----GCAGATCAAGAAACAAAGTCCAGTC 984  
 Db 383 SerAspSerSerSerAspSerAspPheGluProSerGlnAsnHisSerGlnGlyProLeu 402  
 QY 985 TTGTATGAGCAAGATGATCGTTCTCCCAACAGCAATAAATCTTCAAGTTGCCAAGCA 1044  
 Db 403 ArgSerMetValGluAspLeuGlnSerGluGluSerAspGluAspSerSerSerGly 422  
 QY 1045 AGAAGACTGATGCGCATCTTAGGAAACGGTATTGACTAAGGGAAGTGAAGTTAGATTG 1104



423	GlulGuAlaalaaglyLys-----ThrAsnProGlyArgAspSerArgLeu 437	Db
1105	CATTTCCAATTTGAAGGAGAAATAATGCTGGACCACTGACTTA----- 1149	QY
438	SerPheSerAspSerGluSerAspAsnSerAlaAspSerSerLeuProSerArgGluPro 457	Db
1150	-----AATGCCAAGCCATCTGGAAACTTCTTCTAGCCTT 1182	QY
458	ProProProGlnLysProProProAsnSerLysValSerGlyArgSerPro--- 476	Db
1183	AATGTAGATGTAGAGTTCACAG-----CAGCATGGAAAAAGGNTTCTAAATT 1233	QY
477	---GluSerCysSerLysProGluLysIleLeuLysGlyThrTyrAspLysAlaTyr 495	Db
1234	ACGATCATTTTCATG-----AGAATTTCCAAGTCAGAGACAGACAGAAAGAA 1281	QY
496	ThrAspGluLeuValGluLeuHisArgArgLeuMetAlaLeuArgGluArgAsnValLeu 515	Db
1282	CAATGTGAATCAGACATCAAGAACAGAAAGAGATTCAAAATACATCCACCTAAC 1341	QY
516	GlnGlnIleVal-AsnLeuIleGluGluThrGlyHisPheAsnValThrAsnThrThrPh 535	Db
1342	CTCCCTCCAGAGAAGAGTGGCTGGGAACCTCTATTGAGGAATATGAGAAAATGCCTC 1399	QY
535	aAspPheAspLeuPheSerLeuAspGluThrThrValArgLysLeuGlnSerCysLeu 554	Db

## RESULT 13

DSPP\_MOUSE  
 ID DSPP\_MOUSE STANDARD; PRT; 334 AA.  
 AC P97399; O70567;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP-  
 DE 3) [Contains: Dentin phosphoprotein (Dentin phosphophoryn) (DPP);  
 DE Dentin sialoprotein (DSP)].  
 DE DSPP.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_taxid=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Swiss Webster; TISSUE=Molar;  
 RC MEDLINE=97150835; PubMed=8995371;  
 RX MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;  
 RA "Dentin phosphoprotein and dentin sialoprotein are cleavage products  
 RT expressed from a single transcript coded by a gene on human chromosome  
 RT 4. Dentin phosphoprotein DNA sequence determination.";  
 RL J. Biol. Chem. 272:835-842(1997).  
 RN [2]  
 RN SEQUENCE FROM N.A., AND REVISIONS TO C-TERMINUS.  
 RP STRAIN=129/SvJ; TISSUE=Liver;  
 RC MEDLINE=982111969; PubMed=9545272;  
 RX Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B.,  
 RA D'Souza R.N., Kozak C.A., MacDougall M.;  
 RT "Genomic organization, chromosomal mapping, and promoter analysis of  
 RT the mouse dentin sialophosphoprotein (Dsp) gene, which codes for both  
 RT dentin sialoprotein and dentin phosphoprotein.";  
 RL J. Biol. Chem. 273:9457-9464(1998).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=129/SvJ; TISSUE=Liver;  
 RC Sfeir C., Butler S., Lin E., George A., Veis A.;  
 RA "From mouse to zebrafish-dentin matrix proteins genomic  
 RT characterization.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=98055479; PubMed=9395101;  
 RX Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,  
 RA Hotten D., Papagarakis P., Berald A., Butler W.T.;  
 RT "Dentin sialoprotein (DSP) transcripts: developmentally-sustained

expression in odontoblasts and transient expression in  
pre-ameloblasts.";  
Eur. J. Oral Sci. 105:405-413(1997).  
[5]  
TISSUE SPECIFICITY.  
MEDLINE=21096982; PubMed=11175790;  
Xiao S., Yu C., Chou X., Yuan W., Wang Y., Bu L., Fu G., Qian M.,  
Yang J., Shi Y., Hu L., Han B., Huang W., Liu J., Chen Z.,  
Zhao G., Kong X.;  
"Dentinogenesis imperfecta 1 with or without progressive hearing loss  
is associated with distinct mutations in DSPP.";  
Nat. Genet. 27:201-204(2001).  
-I- FUNCTION: DSP may be an important factor in dentinogenesis. DSP  
may bind high amount of calcium and facilitate initial  
mineralization of dentin matrix collagen as well as regulate the  
size and shape of the crystals.  
-I- SUBCELLULAR LOCATION: Secreted (By similarity).  
-I- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and  
transiently in pre-ameloblasts. Found in the inner ear.  
-I- PTM: DSP is glycosylated.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/sib/announcement.html>)  
or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).

ENBL;	U67916;	AAC12787.1;	-
DR	ENBL;	AJ002141;	CRA05208.1; ALT_SEQ.
DR	ENBL;	AF135799;	AAD42781.1; ALT_SEQ.
DR	MGD;	MG1:109172;	Depp.
KW	Biomineralization;	Extracellular matrix;	Signal; Glycoprotein;
KW	phosphorylation;	Sialic acid.	POTENTIAL.
FT	SIGNAL	1	
FT	CHAIN	18	934
FT	CHAIN	18	934
FT	CHAIN	452	934
FT	CHAIN	419	934
FT	DOMAIN	419	934
FT	SITE	479	481
FT	MOD_RES	227	227
FT	MOD_RES	254	254
FT	MOD_RES	279	279
FT	MOD_RES	293	293
FT	MOD_RES	299	299
FT	MOD_RES	314	314
FT	MOD_RES	336	336
FT	MOD_RES	349	349
FT	CARBOHYD	54	54
FT	CARBOHYD	84	84
FT	CARBOHYD	130	130
FT	CARBOHYD	190	190
FT	CARBOHYD	313	313
FT	CARBOHYD	373	373
FT	SEQUENCE	934 AA;	939301 MW; A618789D8A57249A CRC64;

Alignment Scores:	
Pred. No.:	0.0048
Score:	157.50
Percent Similarity:	34.13%
Best Local Similarity:	20.30%
Query Match:	2.33%
Da:	1
Gaps:	18
Indels:	133
Mismatches:	172
Conservative:	64
Matches:	94
Length:	934

US-09-302-812-5 (1-3814) x DSPP\_MOUSE (1-934)

Qy	19	GGTCCAGCATGTCGGGCGCCGGCTGGAG-----CCCTGCACGAAGCGCGCTGG	72
			:::
Db	257	GlyGluGlyAlaGluAlaGlyAspGlyArgGluSerHisAspGlyThrLyselyAlnGly	276
			:::
Qy	73	GGCGCGCT-----GGAACCTTCGGCCGACGTCCTCGGAC	108
			:::

Db 277 GlyGlnSerHisGlyGlyAsnThrAspHisArgGlyGlnSerSerValSerThrGluAsp 296  
 QY 109 TCCCGGAGC-----TTCCCTGGC----- 126  
 Db 297 AspAspSerLysGluGlnGluGlyPheProAsnGlyHisAsnGlyAspAsnSerSerGlu 316  
 QY 127 -----AGCAGAGGGCGTCTTCGAC 147  
 Db 317 GluAsnGlyValGluGluGlyAspSerThrGlnAlaThrGlnAspLysGluLysLeuSer 336  
 QY 148 CCCAGGAGCGCT---CCGTCAGTTCAGGTCCTCCGTCCTCCAGCCTCGCTCTCG 204  
 Db 337 ProLysAspThrArgAspAlaGluGlyGlyIleLeuSerGlnSerGluAlaCysProSer 356  
 QY 205 GGGCGGGCG-----GGACCGCACAGAGGCAACCCACCTCG 240  
 Db 357 GlyLysSerGlnAspGlnGlyGluThrGluGlyProAsnLysGlyAsn----- 373  
 QY 241 TTTGTTTTCAAAACAAAGACTATTACTCTGGATGGATCTAAAGAGCCCAAGACAGCT 300  
 Db 374 -----LysSerIleIleThr-----LysGluSerGlyLysLeuSerGly 386  
 QY 301 GAATCAGAAAGTAAGAAACACATACACAAATTTGACTCCTCATGATGAGTTCTGTGAG 360  
 Db 387 SerLysAspSerAsnGlyHisGlnGlyValGluLeuAspLysArgAsnSerProLysGln 406  
 QY 361 AAAGATAACTTTTACCACATCAAGGTGGAAAAATTTGMAAATGTTCTCAGCTAAATCTT 420  
 Db 407 GlyGluSerAspLysPro----- 412  
 QY 421 GATAAATCACCCACAGAAAGAGTTTACAGTATTTGAACCAACAGCAGCTCGAGTGTG 480  
 Db 413 ---GlnGlyThrAlaGluLysSerAlaAlaHisSerAsn----- 424  
 QY 481 TGCAGTGCAGAAATGACGGAGAGCATGCAAGACAGCTTTTGGCAGTGAGCTCCCGG 540  
 Db 424 ----- 424  
 QY 541 GGGACTCGCTACCAAGCAGCTTAGTAACTGCTAACTGGTTCAGTCACTCCACCACTGAT 600  
 Db 425 -----LeuGlyHisSerArgIleGlySerSerAsnSerAsp 437  
 QY 601 GACCACAGTGACAGATCATGAAGACAGACAGACAAATCAGCAGTTCTTACACCTATA 660  
 Db 438 GlyHisAsp-----SerTyrGluPheAspAspGluSerMetGlnGlyAspAspPro--- 454  
 QY 661 AAATTGCAATCAAAAGCAACAGTACAGAGTGGCGGAGCCAGCAACAGTGAAGTGC 720  
 Db 455 -----LysSerSerAspGluSerAsnGlySerAspGluSerAspThrAsnSerGluSer 472  
 QY 721 AGTGATCTCGCCAGCTCTGTGAAGACTGTACAGCTGTCAACAGAGGAGGTGATGTG 780  
 Db 473 AlaAsnGluSerGlySerArgGlyAspAlaSerTyrThrSerAspGluSerSerAspAsp 492  
 QY 781 CTACAGAGAGTCCTTTGTGATGTTGTCGCCAGGACATGGAACTGGACCAAAAAT 840  
 Db 493 AspAsnAspSerAsp---SerHisAlaGlyGluAspAsp-----SerSerAsp 507  
 QY 841 GACAACTGACTGACGACAGAAAGCCTTAGGTGATTCGCTCCATTTCAGAAAGAA 900  
 Db 508 AspSerSerGlyAspGlyAspSerAspSerAsnGlyAspGly-----AspSerGlu 524  
 QY 901 AGTGAGCTGAGTCCACCAATGATGTAGACACTCGAGAAACAGTTGTCAAGATTCAGAA 960  
 Db 525 SerGluAspLysAspGluSerAspSerSerAspHisAspAsnSer---SerAspSerGlu 543  
 QY 961 CGAGATGAAGAAACAGTCCAGTCTTTGATGACGCAAGATGATGCTCTCCCAACACAGA 1020  
 Db 544 SerLysSerAspSerSerAspAspAspSerSerSerSerSerSerSerSerSerSer 563  
 QY 1021 AATAAATTTCAAGTTGCCACAGACAGAGTGTGCGGATCTTAGGAACGGTATTG 1080  
 Db 564 SerSerAspSerSerAspSerSerSerSerSerSerSerSerSerSerSerSerSer 576

QY 1081 ACTAAGCGAAGTGAAGTTAGATTGCATTTCCAAATTTGAAGGAGAAAATAATCTGGGACC 1140  
 Db 577 SerAspSerSerAsp-----SerSerAspSerAsnSerSerSer 589  
 QY 1141 AGTGACTTTAAATGCGCAAGCCATCTCGAAACTCTTCTAGCCTTAAATGATAGAGTGAAGT 1200  
 Db 590 AspSerSerAspSerSerGlySerSerAspSerSerAspSerSerAspThrCysAspSer 609  
 QY 1201 TCCAGCAGCATGGAAAAGGATCTAAATTTACAGATCATTTTATGAGAATTTCCAAAG 1260  
 Db 610 SerAspSerSerAspSerSerAspSerSerAspSerSerAspSerSerSerSerAsp 629  
 QY 1261 TCAGAGGAGC 1269  
 Db 630 SerSerAsp 632  
 RESULT 14  
 ATRX CAEEL  
 ID ATRX CAEEL STANDARD; PRT; 1359 AA.  
 AC Q9U7E0; O02061;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).  
 GN XNP-1 OR B0041.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=62319;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=9365296; PubMed=10433961;  
 RA Villard L., Fontes M., Ewbank J.J.;  
 RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to  
 the human XNP/ATR-X gene.";  
 RL Gene 236:13-19 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RC Fulton R., Wohldmann P.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies  
 gene expression by affecting chromatin (Potential).  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AF134186; AAD55361.1; -.  
 CC EMBL; AF000196; AAC24256.1; -.  
 CC PIR; T34036; T34036.  
 CC WormPep; B0041.7; CE17314.  
 CC InterPro; IPR001410; DEAD.  
 CC InterPro; IPR001650; Helicase\_C.  
 CC InterPro; IPR000330; SNF2\_N.  
 CC Pfam; PF00271; Helicase\_C; 1.  
 CC Pfam; PF00176; SNF2\_N; 1.  
 CC SMART; SM00487; DEXDC; 1.  
 CC SMART; SM00490; HELIC; 1.  
 CC PROSITE; PS00690; DEAD\_ATP\_HELICASE; FALSE NEG.  
 CC DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;  
 KW NP BIND. 496 503 ATP (POTENTIAL).  
 FT SITE 636 639 DEAD BOX.  
 FT DOMAIN 67 70 POLY-ASP.  
 FT DOMAIN 266 272 POLY-GLU.





Alignment Scores:	
Pred. No.:	0.00589
Score:	157.50
Length:	2211
Matches:	135
Percent Similarity:	32.87%
Conservative:	102
Best Local Similarity:	18.72%
Mismatches:	295
Query Match:	2.33%
Indels:	189

Db 1271 SerProAspLeuGlyGlnThrAlaLeuSerProAspProGlyGlnThrAlaLeuSerPro 1290

```
Db 977 AsnSerHisGlyLysGlnSerGlyHisProThrPheLeuValThrArgArgLysProLeu 996
Qy 988 ---GATGAGCAAGATGATCGTCTCCCAACAGCAAAATAAATCTTCAAGTTGCCAAGCA 1044
Db 997 GlnAspArgGlnAspArgAsnSerArgLeuLysGluGlyLeuProLeuLeuArgThr 1016
Qy 1045 AGAAGAGCTGATGCGATCTTAAAGAAACGGTATTGACTAAGGAAGTGAAGTTAGATTG 1104
Db 1017 ArgArgLysLysGluGluLysProAlaThrHisValProLeuSerProArgSerPhe 1036
Qy 1105 CATTTCCAAATTTGAAGAGAAATAATCTCGGACCACTGACTTAATGCAAGCCATCT 1164
Db 1037 His---ProLeuArgGlyGluValAsnAlaSerPheSerAspArgArgHisAsnHisSer 1055
Qy 1165 -----GGAACTCTTCTAGCTTAAATGATAGTGTAGAGTTCC----- 1203
Db 1056 LeuLeuLeuHisAlaSerAsnGluThrSerLeuSerIleAspLeuAsnGlnThrPhePro 1075
Qy 1204 AAGCAGCATGGAAAAAGGATTCTAAATTTACAGATCATTTTCATGAGAATTTCCAAGTCA 1263
Db 1076 SerMetAsnLeuSerLeuAlaAlaSerLeuProAspHis----- 1088
Qy 1264 GAGGACAGAGAAAGAAACAATGTCAAGTTCAGACATCAAAAGACAGAAAGAGATTCCA 1323
Db 1089 -----AspGlnThrSerProAsnAspThrThrSerGlnThrSerSerPro 1103
Qy 1324 AAATACATCCACCTAACTCCTCCAGAGAAAGTGGCTGGGAACCTCTATTGAGGAA 1383
Db 1104 ProAspLeuTyProThrValSerProGluGluHisTyrGlnIlePheProIleGlnAsp 1123
Qy 1384 ATGAGAAAAATCGCTCGGTGGGATCCATTTGCCCTTCTTAAAGA---CCATCTGCAAGT 1440
Db 1124 -----SerAspProThrHisSerThrThrAlaProSerAsnArgSerProAspProThr 1141
Qy 1441 CACACAGTACTGTGT-----CGGTAGACCTTCTGAGAGACAGAGAGGTCCGAAA 1491
Db 1142 HisSerThrThrAlaProSerAsnArgSerProProThrGlnProSerGlnIleProAsn 1161
Qy 1492 CTTTTTCCAAACATACAAAGATTGTTGGGATACAAACATGTGMAATGCTGTGTTGG 1551
Db 1162 -----TyrAspLeuArgAsnArgAlaIleProThrAsp 1172
Qy 1552 GAACAAAACCTGTACCTGTGGAAGATGAGATGTGAGCGAATCTCGAGGAGTAGGTGG 1611
Db 1173 ValSerGlnIlePhePro-----SerLeuGluLeu 1182
Qy 1612 GAGCTCATTACAGACTGCACTTCTCAACAAATTCACAGACCCCACTTGAAGGATGCG 1671
Db 1183 GluValTrpGlnThrAlaThrSerLeuAspLeuSerGlnPro----- 1196
Qy 1672 ATCTGGAATACAAATGTCATATTCTAAGAAATGGGACTTTACAGCTTTGGTTGATTTC 1731
Db 1197 -----SerIleSerProAspLeu 1202
Qy 1732 TGGGATAGGTACTTGAGAGACAGAGCCCAACATTTATATCAGTCCATTTTACTGAC 1791
Db 1203 GlyGlnMetAlaLeuSerProAspProGlyGln-----GluSerLeuSerProAsp 1219
Qy 1792 ATGGTGAATAATGCACTCTGTCTGCCAATATTTGCAACCCAGCCCAATA---CCACTCTG 1848
Db 1220 LeuGlyGlnThrSerLeuSer---ProAspLeuSerGlnGluSerLeuSerProAspLeu 1238
Qy 1849 AAACAGAGATGAATCATTTCTGTCACGATGTACAGGAACAGATCGCCAGTCTTTTAGCT 1908
Db 1239 GlyGlnThrAla---LeuSerProAspProSerGlnGluSerLeuSerProAspLeuGly 1257
Qy 1909 AATGCTTTCTCTGCACTTTCGCCAGCATGTCACAGATGAATCGAGATTTCTAGT 1968
Db 1258 GlnThrAlaLeuSerProAspPro-----SerGlnGluSerLeu 1270
Qy 1969 TACCCAGACATTAATCTCAATCGGTGTTT-----GAAGGAGTTCATCAAGGAACCA 2022
```

Search completed: May 26, 2004, 16:55:49  
Job time : 221.908 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:13:47 ; Search time 178.048 Seconds  
(without alignments)  
13517.524 Million cell updates/sec

Title: US-09-302-812-5  
Perfect score: 6773  
Sequence: 1 ggggagctgtgtgcggg.....aatcattgtcagaaaaaa 3814

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Maximum Match 0%

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -MODE=frames+...  
-Q/cgn2\_1/USPTO.spool\_p/6333148/runat.26052004.150053.5676/app.query.fasta\_1.12437  
-DB=SPTRMBL\_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=6333148@cgn\_1\_1.640@runat.26052004.150053.5676 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

SPTRMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rviris:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5164	76.2	968	11	O88622 mus musculus

	2	5098.5	75.3	961	11	Q80YQ6
3	4869.5	71.9	920	11	Q8CB72	Q80YQ6 mus musculus
4	4749	70.1	972	11	Q9QYM2	Q8CB72 mus musculus
5	4456	65.8	976	4	Q86W56	Q9QYM2 rattus norv
6	4451	65.7	976	4	Q72742	Q86W56 homo sapien
7	4395	64.9	976	4	Q9Y4W7	Q72742 homo sapien
8	4332.5	64.0	977	6	O02776	Q9Y4W7 homo sapien
9	1074.5	15.9	768	5	Q46043	O02776 bos taurus
10	1069.5	15.8	723	5	Q960N8	Q46043 drosophila
11	568	8.4	548	10	Q9SKB3	Q960N8 drosophila
12	528	7.8	522	10	Q8VYAL	Q9SKB3 arabidopsis
13	458	6.8	485	5	Q9NSL4	Q8VYAL arabidopsis
14	451	6.7	781	5	Q867X0	Q9NSL4 caenorhabdi
15	450.5	6.7	764	5	Q19637	Q867X0 caenorhabdi
16	328.5	4.9	368	5	Q86G14	Q19637 caenorhabdi
17	328	4.8	364	10	Q9SKB4	Q86G14 toxoplasma
18	262.5	3.9	747	4	Q86V50	Q9SKB4 arabidopsis
19	190	2.8	931	5	Q9VF92	Q86V50 homo sapien
20	188.5	2.8	930	5	Q81HH1	Q9VF92 drosophila
21	183	2.7	1194	11	Q8C0H4	Q81HH1 drosophila
22	183	2.7	1194	11	Q8BM10	Q8C0H4 mus musculus
23	182	2.7	571	5	Q8MNT7	Q8BM10 mus musculus
24	179.5	2.7	1193	11	Q8BTQ4	Q8MNT7 trichinella
25	177.5	2.6	3166	5	Q9W3Z0	Q8BTQ4 mus musculus
26	177	2.6	5322	5	Q9VPL9	Q9W3Z0 drosophila
27	176.5	2.6	888	12	Q8UKL5	Q9VPL9 drosophila
28	176	2.6	8081	5	Q72120	Q8UKL5 heliothis z
29	173	2.6	1537	4	Q96NW7	Q72120 caenorhabdi
30	172	2.5	508	12	Q9E234	Q96NW7 homo sapien
31	169	2.5	700	4	Q98UV3	Q9E234 helicoverpa
32	167	2.5	510	12	Q99GU9	Q98UV3 homo sapien
33	167	2.5	1169	16	Q7ZAJ1	Q99GU9 helicoverpa
34	167	2.5	1758	4	Q7ZAE8	Q7ZAJ1 staphylococ
35	166	2.5	1616	4	Q96PH3	Q7ZAE8 homo sapien
36	166	2.5	1792	4	Q7Z6E9	Q96PH3 homo sapien
37	165	2.4	498	5	Q8MTN8	Q7Z6E9 homo sapien
38	165	2.4	1192	5	Q96127	Q8MTN8 trichinella
39	164	2.4	766	11	Q8C1D8	Q96127 plasmodium
40	163	2.4	532	10	Q9T0E9	Q8C1D8 mus musculus
41	163	2.4	691	4	Q96QF7	Q9T0E9 arabidopsis
42	163	2.4	2392	5	Q810K1	Q96QF7 homo sapien
43	163	2.4	2407	5	Q7YSH5	Q810K1 caenorhabdi
44	163	2.4	2677	4	Q7Z333	Q7YSH5 caenorhabdi
45	162.5	2.4	1508	5	Q86J39	Q7Z333 homo sapien
						Q86J39 dictyosteli

## ALIGNMENTS

RESULT 1

O88622

ID O88622 PRELIMINARY; PRT; 968 AA.

AC O88622

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE Poly(ADP-ribose) glycohydrolase.

GN PARG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]\_SEQUENCE FROM N.A.

RP MEDLINE=99380098; PubMed=10449915;

RX Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;

RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to

RT human chromosome 10q11.23 and mouse chromosome 14B by in situ

RT hybridization."

RL Cytogenet. Cell Genet. 85:269-270 (1999).

DR EMBL; AF079557; AAC28735.1; -.

DR MGD; MGI:1347094; Parg.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR InterPro; IPR007724; PARG.

DR Pfam; PF05028; PARG; 1.  
KW Hydrolase. 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 968  
Score: 5164.00 Matches: 968  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 76.24% Indels: 0  
DB: 11 Gaps: 0  
US-09-302-812-5 (1-3814) x 088622 (1-968)  
QY 28 ATGAGTGGGGCCCGCTGGAGCCCTGACGAAAGCGGCTGGGGCGGCTGGAACT 87  
DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLysAlaArgTrpGlyAlaGlyThr 20  
QY 88 TCTGGCGCGGACTGCTCGGACTCCGGAGCTTCCCTGGCAGGCGAGGGCGTGTCTCGAC 147  
DB 21 SerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeuAsp 40  
QY 148 CCCAAGACCGCTCCCGTCCAGTTCAGGTCCTCCGTCCTCGCAGCGCTCGCTCGGG 207  
DB 41 ProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSerGly 60  
QY 208 CGGGCGGGACCGCACAGAGCGCAACGCCACCTCGTTGTTTTCAAACAAAGACTTACT 267  
DB 61 ArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIleThr 80  
QY 268 ACTTGATGATGATTAAGGACCCCAACAGCAGCTGAATCAGAAAGTAAAGAAAAACAACAT 327  
DB 81 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGluSerLysGluAsnAsn 100  
QY 328 ACAAGATTGACTCCATGATGAGTTCGTGCGAAGAGATTAACCTTTACCCACATAAGGTG 387  
DB 101 ThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyrProHisLysVal 120  
QY 388 GAAAAATTCGAAATGTTCTCAGCTAAATCTTGATTAATCACCACGAAAGAGTTCA 447  
DB 121 GluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThrGluLysSer 140  
QY 448 CAGTATTGAAACCAACAGCAGACTGCGAGTGTGTGCAAGTGGCAGAGTAAAGGGAAGCAT 507  
DB 141 GlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsnGluLysHis 160  
QY 508 GCGAAGACGCTTTGGCAAGTGAAGTCCCGCGGAGCTCCGTACAAAGCAGCTTAGT 567  
DB 161 AlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuProLysGlnLeuSer 180  
QY 568 AATGCTAACATTGCTCAGTACCCACCTGATGATGACCAAGTACACAGATCATGAGAA 627  
DB 181 AsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThrAspHisGluGlu 200  
QY 628 GACAGAGACAATCAGCAGATTTCTTACACCTATATAAACTTGCAATACAAAGCCAAACAGTA 687  
DB 201 AspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThrLysProThrVal 220  
QY 688 GGAGATGGGCGCCAGAACCAACTGTAAGTGAAGTCTCGCAGTCTCGGAGTCTGGAAGAC 747  
DB 221 GlyAspGlyGlnAlaArgSerAsnCysLysCysSerGlySerArgGlnSerValLysAsp 240  
QY 748 TGTACAGGCTGCAACAGGAGGAGGTGGATGCTACCTACGAGAGTCTTTGTGTCAGATGTT 807  
DB 241 CysThrGlyCysGlnGlnGluGluValAspValLeuProGlnSerProLeuSerAspVal 260  
QY 808 GGTGCGCGAGACATTGGAATCGGACCAAAAAATGACAAACAAATTTGACTGGACAAAGAAC 867  
DB 261 GlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGluSer 280  
QY 868 AGCTAGGTGATTCGCTCCATTGTGAAAGAAAGATGAGCTCGATGACCAATGGATGTA 927  
DB 281 SerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAspVal 300

QY 928 GACAACTCGAANAACAGTTGTCAAGATTCAAGACGACGATCAAGAAACAAAGTCCAGTCTTT 987  
DB 301 AspAsnSerArgAsnSerCysGlnAspSerGluAlaAspGluGluThrSerProValPhe 320  
QY 988 GATGACCAAGATCATCGTTCTCCCAACACAGCAAAATAAACTTTCAAGTTCGCCAAGCAAGA 1047  
DB 321 AspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerCysGlnAlaArg 340  
QY 1048 GAAGCTGATGGCGATCTTTAGAAACCGTATTTGACTTAAGGAAAGTGAAGTTAGATGTCAT 1107  
DB 341 GluAlaAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGluValArgLeuHis 360  
QY 1108 TTCCAAATTTCAAGGAGAAAATAATGCTGGGACAGTCACTTAATGCCAGCCATCTGGA 1167  
DB 361 PheGlnPheGluGlyGluAsnAsnAlaGlyThrSerAspLeuAsnAlaLysProSerGly 380  
QY 1168 AACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCGAAGTTCGAAGCAGCATGAAAAAGGATCT 1227  
DB 381 AsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAspSer 400  
QY 1228 AAAATTACAGATCATTTCTAGAAATTCAAAGTCAAGACAGACAGAAAGAAACAATGT 1287  
DB 401 LysIleThrAspHisPheMetArgIleSerLysSerGluAspArgArgLysGluGlnCys 420  
QY 1288 GAAGTCAGACATCAAGAAACAGAAAGAGATTCCAAAATACATCCACCTAACCTCCCT 1347  
DB 421 GluValArgHisGlnArgThrGluArgLysIleProLysTyrIleProAsnLeuPro 440  
QY 1348 CCAGAGAAGAAGTGGCTGGGAACTCCTATTGAGGAATCAGAAAAATGCTCGGTGTGGG 1407  
DB 441 ProGluLysLysTrpLeuGlyThrProIleGluGluMetArgLysMetProArgCysGly 460  
QY 1408 ATCCATTGCTTCCCTTAAGACCATCTGCAAGTCAACAGTGAAGTCTGTTGGGTAGACCTT 1467  
DB 461 IleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAspLeu 480  
QY 1468 CTGAGAGCAGGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTGTGGGATAAC 1527  
DB 481 LeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAspLeuTrpAspAsn 500  
QY 1528 AAACATGTGAAATGCTTTGTTTCGGAAACAAACTTGTACCTGTGGAAGATGAGATGGT 1587  
DB 501 LysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGluAspGluAsnGly 520  
QY 1588 GAGCGAAGTCCGAGGATAGGTGGAGCTCATTCAGACTGCATCTTCTCAACAATTCACA 1647  
DB 521 GluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuAsnLysPheThr 540  
QY 1648 CGACCCCAACTGTGAAGGATCGGATCTGGAATCAATGTGGCATATTCTAAGAATGG 1707  
DB 541 ArgProGluAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyrSerLysLysTrp 560  
QY 1708 GACTTTACAGCTTGGTGTGATTTCTGGATAAGTACTTGAAGACGACAGGCCCCAACAT 1767  
DB 561 AspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAlaGluAlaGlnHis 580  
QY 1768 TTATATCAGTCCATTTTACTGACATGGTGAATTTGCACTCTGTCTGCAATATTTCG 1827  
DB 581 LeuTyrGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIleCys 600  
QY 1828 ACCAGCCCAATCACTCCTCGAAACAGAAAGATGAATCATTTGTCTGACGATGTCCACAGGAA 1887  
DB 601 ThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGlnGlu 620  
QY 1888 CAGATGCCAGTCTTTTAGCTAATGCTTTCTGCACTTTCCCGACGGAATGCCAAG 1947  
DB 621 GlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAlaLys 640  
QY 1948 ATGAAATCGAGTATCTAGTTTACCCAGACATTAACCTCAATCGGTGTGTTTGAAGACGT 2007  
DB 641 MetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeuPheGluGlyArg 660





745 GACTGTACAGGCTGTCACACAGAGGAGGTGGATGCTCTACAGAGAGTCTTTGTGCAGAT 804  
241 AspCysThrGlyCysGlnGlnGluGluValAspValLeuProGluSerProLeuSerAsp 260  
805 GTTGGTCCGAGACATTGGAACTGGACCAAAAATGACAAACAAATTTGACTGCACAAGAA 864  
261 ValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsnLysLeuThrGlyGlnGlu 280  
865 AGCAGCCTAGTGATTCGCTCCTCATTTGAGAAAGAAAGTGAGCTGAGTCCCAATGGAT 924  
281 SerSerLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAsp 300  
925 GTAGACAACTCGAGAACAGTTCCTCAAGATTCAAGACGAGATCAAGAAACAAGTCCAGTC 984  
301 ValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGluThrSerProVal 320  
985 TTTGATGAGCAAGATGATCGTCTCCCAACAGCAAAATAAACTTTCAAGTTGCCAAGCA 1044  
321 PheAspGluGlnAspAspArgSerSerGlnThrAlaAsnLysLeuSerSerCysGlnAla 340  
1045 AGAAGACTGATCGCGATCTTAGAAACGCTATTTCAGTAAAGGAGAGTGAAGTTAGATTG 1104  
341 ArgGluAlaAspGlyAspLeuArgLysArgLysLeuThrLysGlySerGluValArgLeu 360  
1105 CATTTCCAAATTTCAAGGAGAAAAATAATGCTGGACCAAGTGAATTAATGCCAAGCCATCT 1164  
361 HisPheGlnPheGluGlyGluAsnAsnValGlyThrSerAspLeuAsnAlaLysProSer 380  
1165 GGAACCTCTTCTAGCCTTAATGTAGAGTGTAGAGTTCCAGAGCAGCAGTGAAGAAAGGAT 1224  
381 GlyAsnSerSerSerLeuAsnValGluCysArgSerSerLysGlnHisGlyLysArgAsp 400  
1225 TCTAAATTTACAGTCAATTTATGAGCAATTTCCAGTTCAGAGCAGACAGAAAGAACAA 1284  
401 SerLysIleThrAspHisPheMetArgLysSerLysSerGluAspArgArgLysGluGln 420  
1285 TGTGAAGTCAGACATCAAGAAACAGAAAGAAAGATTCCAAATACATCCCAACCTTAACCTC 1344  
421 CysGluValArgHisGlnArgThrGluArgLysIleProLysThrIleProAsnLeu 440  
1345 CCTCCAGAGAAAGTGGCTGGAACTCTTATTGAGGAATGAGAAATGCTCGTGT 1404  
441 ProProGluLysLysTrpLeuGlyThrProIleGluLysMetArgLysMetProArgCys 460  
1405 GCGATCCATTGCTCTTAAAGACCATCTGCAAGTCACAGTGAAGTCTGCTGGTAGAC 1464  
461 GlyIleHisLeuProSerLeuArgProSerAlaSerHisThrValThrValArgValAsp 480  
1465 CTTCTGAGAGCAGAGAGGTTCCGAAACCTTTTCCAAACACATTACAAAGATTTGTGGAT 1524  
481 LeuLeuArgAlaGlyGluValProLysProPheProThrHisLysAspLeuTrpAsp 500  
1525 AACAAATGTGAATTCCTTGTTCGGAACAAACTGTACCTGTGGAGATGAGAAT 1584  
501 AsnLysHisValLysMetProCysSerGluGlnAsnLeuTrpProValGluAspGluAsn 520  
1585 GGTGAGCGAACTGAGAGAGTGTGGAGTCTATTGAGCTGACATTCCTCAACAAATTC 1644  
521 GlyGluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeuAsnLysPhe 540  
1645 ACAGACCCCAAGACTTGAAGGATCGATTCGAAATACAAATGAGTGGCATATTTAAGAAA 1704  
541 ThrArgProGlnAsnLeuLysAspAlaIleLeuLysLysTrpAsnValAlaLysSerLysLys 560  
1705 TGGACATTTACAGCTTTGGTTGATTTCTGGATTAAGTACTTGAAGACGAGCGGCCAA 1764  
561 TrpAspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluValAlaGluAlaGln 580  
1765 CATTTATATCATGCTTTTACCTGACATGTTGAAATTTGACATCTGTCTGCCAAATATT 1824  
581 HisLeuTrpGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeuProAsnIle 600  
1825 TGCACCCAGCCCAATACCTCTCGAAACAGAAAGATGAATCATTTCTGTCCAGATGTACAG 1884

601 CysThrGlnProIleProLeuLeuLysGlnLysMetAsnHisSerValThrMetSerGln 620  
1885 GAACAGATCGCAGTCTTTAGCTAATGCTTTCTGACATTTCCCGACGAGGAATGCC 1944  
621 GluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArgArgAsnAla 640  
1945 AAGATGAATCCGAGTATTCTAGTTTACCAGACATTAACTCAATCGTGTGTTGAAGA 2004  
641 LysMetLysSerGluTrpSerSerTrpProAspIleAsnPheAsnArgLeuPheGluGly 660  
2005 CTTTCATCAAGAAACACGAAAAACCTCTTCTGCTACTTTTCAGAGAGTCACA 2064  
661 ArgSerSerArgLysProGluLysLeuLysThrLeuPheCysTrpPheArgArgValThr 680  
2065 GAGAAAAACCTTACAGGATTTGTGACATTTACAAGACAGAGTCTTCAAGATTTTCAGAA 2124  
681 GluLysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAspPheProGlu 700  
2125 TGGGAAAGGTGTGAAAGCCTCTGACACGCTTACAGTCACTTACGAGGAGTACATAGAA 2184  
701 TrpGluArgCysGluLysProLeuThrArgLeuHisValThrTrpGluGlyThrIleGlu 720  
2185 GGCACGCGCGAGGCTGCTACAGGTGGATTTTGCAAATCGTTTCTGAGGTGTGTG 2244  
721 GlyAsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGlyGlyVal 740  
2245 ACTGGTCCGGCTTGTCAAGAAAGAAATCAGATTTTAAATCAATCTGAATTTGATGTT 2304  
741 ThrGlyAlaGlyLeuValGlnGluGluIleArgPheLeuIleAsnProGluLeuIleVal 760  
2305 TCACGCTCTTCACTGAGGTGTGATCAATAGTGTCTTATATTCACAGTACTGAA 2364  
761 SerArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleIleThrGlyThrGlu 780  
2365 CAGTACAGTGAATACACAGCTATGCTGAACTTATCGTTGGCCGAGCCATGAAGAT 2424  
781 GlnTrpSerGluTrpThrGlyTrpAlaGluThrTrpArgTrpAlaArgSerHisGluAsp 800  
2425 GGGAGTGAAGAGCATTGGACGCGCTGCAGCGAGATCGTTCGCAATGACGACCTT 2484  
801 GlySerGluLysAspAspTrpGlnArgArgCysThrGluIleValAlaIleAspAlaLeu 820  
2485 CACTTCAGACGCTTACCTCGATCAGTTTGTGCTGAGAAAGTGAAGTGAAGTGAAGTGAAG 2544  
821 HisPheArgArgTrpLeuAspGlnPheValProGluLysValArgArgGluLeuAsnLys 840  
2545 GCTTACTGCGGATTCCTCGTCTCGAGTTCCTTCTGAAATCTTCTGAGTGGCCACG 2604  
841 AlaTrpCysGlyPheLeuArgProGlyValProSerGluAsnLeuSerAlaValAlaThr 860  
2605 GGAACCTGGGGCTGTGCTGCTTTGGGGGTGACGCTAGATTAAAGCCTTAAATACAGATC 2664  
861 GlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeuIleGlnIle 880  
2665 CTGGAGCTCTCGGCTGAACTGAGTGGTGTATTTCACCTTGGGACTCAGAGTTG 2724  
881 LeuAlaAlaAlaAlaGluArgAspValValTrpPheThrPheGlyAspSerGluLeu 900  
2725 ATGAGACATTTTACAGCATGACATTTCTTCTTACCGAGAGAGTGTGGATGTTGGAAGA 2784  
901 MetArgAspIleTrpSerMetHisThrPheLeuThrGluArgLysLeuAspValGlyLys 920  
2785 GTGTACAAGTTATTGCTTAGATACATACTCAATGAAGATGCAAGAACTGTTCCACCCCTGGA 2844  
921 ValTrpLysLeuLeuLeuArgTrpAsnGluGluCysArgAsnCysSerThrProGly 940  
2845 CCGACATCAAGCTTTATCCATTTCATATACCATGCTGCTTCAAGTCAAGTGCAGAGACCAT 2904  
941 ProAspIleLysLeuTrpProPheIleTrpHisAlaValGluSerSerAlaGluThrThr 960  
2905 GAC 2907





141 ThrProTyrLeuAsnGlnGlnThrAlaGlyValCysLysTrpHisSerAlaGlyGlu 160 Db  
505 CATGCAGAAACAGCTTTTGGCAAGTAGCCTCCCGGGGAGCTCGCTACCAAGCAGCTT 564 Qy  
161 ArgAlaGlnGlnLeuSerAlaSerGluProSerAlaValThrGlnAlaProLysGlnLeu 180 Db  
565 AGTAATGCTTAACATTTGCTCAGTCCACCCACACTGATGACACAGTACACAGATCAGAA 624 Qy  
181 SerAsnAlaAsnIleAspGlnSerProProThrAspGlyHisSerAspThrAspHisGlu 200 Db  
625 GAACACAGACACATCAGCAGTTCTTACACTATAAACTTGCATAATACAAACCCAAACA 684 Qy  
201 GluAspArgAspAsnGlnGlnPheLeuThrProValLysLeuAlaAsnAlaLysGlnThr 220 Db  
685 GTAGCAGATGGCCAGGCCAGAACCACTGTAAGTGCAGTGGATCTCCGCCAGTCTGTGAAA 744 Qy  
221 ValGlyAspGlyGlnAlaArgSerAsnCysLysCysSerAlaSerCysGlnCysGlyGln 240 Db  
745 GACTGTACAGCTGTCAACAGAGAGGTGATGCTGTACACAGAGTCCCTTTGTCAGAT 804 Qy  
241 AspCysAlaGlyCysGlnArgGluGluAlaAspValIleProGluSerProLeuSerAsp 260 Db  
805 GTTGGTCCGAGGACATTTGGAACCTGGACCAAAAAATGACAAATAATGACTGGCAAGAA 864 Qy  
261 ValGlyAlaGluAspIleGlyThrGlySerLysAsnAspAsnLysLeuThrGlyGlnGlu 280 Db  
865 AGCAGCTAGTGATTCGCTCCATTGAGAAAGAGTAGCCGTGAGTACCAATGGAT 924 Qy  
281 SerGlyLeuGlyAspSerProProPheGluLysGluSerGluProGluSerProMetAsp 300 Db  
925 GTAGCAACTCGAAGAACAGTTGTCAGATTCAGAACGAGATGAAGAACAGTCCAGTC 984 Qy  
301 ValAspAsnSerLysThrSerCysGlnAspSerGluAlaAspGluGluAlaSerProVal 320 Db  
985 TTTGATGAG-----CAAGATGATCGTCTCTCCCAACAGCAAAATAAACTTTCAAGT 1035 Qy  
321 PheAspGluGlnAspAspGlnAspArgSerSerGlnThrAlaAsnLysLeuSerSer 340 Db  
1036 TGCCAGCAAGAGAGCTGATGGCGATCTTAGAAACCGTATTGCACTAAGGGAGTGAA 1095 Qy  
341 ArgGlnAlaArgGluValAspGlyAspLeuArgLysArgTyrLeuThrLysGlySerGlu 360 Db  
1096 GTTAGATTGCATTTCCATTTGAAGGAGAAATAATGCTGGACCCAGTCAATTAATGCC 1155 Qy  
361 IleArgLeuHisPheGlnPheGlnGlySerAsnAlaGlyThrSerAspLeuAsnAla 380 Db  
1156 AAGCCATCTGGAAACTCTTCTAGCCTTAATGTAGAGTGTAGAAGTTCCAAGCAGCATGGA 1215 Qy  
381 LysProSerGlyAsnSerSerSerLeuAsnValAspGlyArgSerSerLysGlnHisGly 400 Db  
1216 AAAAGGGATCTTAAATTAACAGATCATTTTCATGAGAAATTCAGTCCAGTACAGAGCAGA 1275 Qy  
401 LysArgAspSerLysIleThrAspHisPheValArgIleProLysSerGluAspLysArg 420 Db  
1276 AAAGAACATGTAGTACAGATCAAGAACAGAAAGAGATTCCAAAATACATCCCA 1335 Qy  
421 LysGluGlnCysGluValArgHisGlnArgAlaGluArgLysIleProLysTyrValPro 440 Db  
1336 CCTAACCTCCCTCCAGAGAAAGTGGTGGGAATCTCTATTGAGGAAATGAGAAAAATG 1395 Qy  
441 ProAsnLeuProProAspLysLysTrpLeuGlyThrProIleGluGluMetArgLysMet 460 Db  
1396 CTCTGGTGGGATCCATTTGCTTCCCTTAAGACCTCTGCAAGTCCAGTACAGTGTCTGTT 1455 Qy  
461 ProArgCysGlyValArgLeuProLeuArgProSerAlaSerHisThrValThrVal 480 Db  
1456 CGGGTAGCTTCTCAGACAGAGAGTTCGAAACCTTTTCCAAACATTAACAAAGAT 1515 Qy  
481 ArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThrHisTyrLysAsp 500 Db  
1516 TTGTGGGATAACAAACATGTGAAATGCTTGTTCGAAACAAACATGTACCTGTGGAA 1575 Qy  
501 LeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeuTyrProValGlu 520 Db

1576 GATGAGAATCGTGCAGCAACTCGAGGAGTAGTGGAGCTCATTCAGACTGCACTTCTC 1635 Qy  
521 AspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGlnThrAlaLeuLeu 540 Db  
1636 AACAAATTCACAGACCCCAAGAACTTGAAGGATGCGATTTCTGAATAATCAATGTGGCATAT 1695 Qy  
541 AsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyrAsnValAlaTyr 560 Db  
1696 TCTAAGAAATGGGACTTTACAGCTTTGTTGATTCTTGGGATAGGTACTTGAAGAGCA 1755 Qy  
561 SerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLysValLeuGluGluAla 580 Db  
1756 GAGSCCAACATTTATATCATCAGTCCATTTTACCTCAGATGGTGAATAATGCACTCTGTCTG 1815 Qy  
581 GluAlaGlnHisLeuLysGlnSerIleLeuProAspMetValLysIleAlaLeuCysLeu 600 Db  
1816 CCAATAATTTTGCACCCAGCCCAATACCACTCTGTAACAGAGATGAATCATCTGTCAAG 1875 Qy  
601 ProAsnIleCysThrGlnProIleProLeuLysGlnLysMetAsnHisSerValThr 620 Db  
1876 ATGTCACAGGAAACAGATCGCCAGTCTTTTACTTAATGCTTCTTCTGCACATTTCCCGCA 1935 Qy  
621 MetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCysThrPheProArg 640 Db  
1936 CGGAATCCCAAGATGAATCGAGTATTCTAGTTTACCCAGACATTAACCTCAATCGGTTG 1995 Qy  
641 ArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsnPheAsnArgLeu 660 Db  
1996 TTTGAAGCAGCTTCATCAAGGAAACCCAGAAACCTGAAACCACTCTTCTGCTACTTTCGA 2055 Qy  
661 PheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPheCysTyrPheArg 680 Db  
2056 AGACTCAGAGAAACCACTACAGAGTTCGTGATTTTACAGACAGAGTCTTCAAGAT 2115 Qy  
681 ArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGlnSerLeuGluAsp 700 Db  
2116 TTTCCAGAAATGGGAAAGGTGTGAAAGCTCTGACACGCTTACACGTCACCTTACAGGGT 2175 Qy  
701 PheProGluTrpGluArgCysAspLysProLeuThrArgLeuHisValThrTyrGluGly 720 Db  
2176 ACCATAGAGCAACCGCCGAGGATGCTACAGGTGGATTTTGCMAATCGTTTGTTCGA 2235 Qy  
721 ThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAlaAsnArgPheValGly 740 Db  
2236 GGTGCTGACTGCTGCGGACTTGTACAAGAGAAATCAGATTTTAAATCAATCCTGAA 2295 Qy  
741 GlyGlyValThrGlyAlaGlyLeuValGlnGluIleArgPheLeuIleAsnProGlu 760 Db  
2296 TTGATTGTTTCCAGCGCTGTTCACTGAGGTGCTGATCAATGAGTGTCTTATTATACA 2355 Qy  
761 LeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGluCysLeuIleThr 780 Db  
2356 GGTACTGAACAGTACAGTGAATACACAGGCTATCTCTGAACTTATCTGTTGGCCGCAAGC 2415 Qy  
781 GlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArgTrpAlaArgSer 800 Db  
2416 CATGAGATGGGAGTGAAGAGGATTTGGCAGCGCGCTGCACGAGATCGTTCCTCAT 2475 Qy  
801 HisGluAspGlySerGluLysAspAspTrpGlnArgCysCysThrGluIleValAlaIle 820 Db  
2476 GACGCACTTCACTTCCAGACGCTACCTCGATCAGTTTGTGCTGAGAAAGTAGAGCTGAG 2535 Qy  
821 AspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLysValArgArgGlu 840 Db  
2536 CTTAAACAGGCTTACTGCGGATTCCTCGTCTCGAGTTCTCTGAAATCTTCTTCTGCA 2595 Qy  
841 LeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProProGluAsnLeuSerAla 860 Db  
2596 GTGGCCACGGGAAACCTGGCTGTGCTGCTTGGGGGTGACGCTAGATTTAAAGCCTTA 2655 Qy  
861 ValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArgLeuLysAlaLeu 880 Db







KW Hypothetical protein.  
SQ SEQUENCE 976 AA; 111090 MW; 47C8EE1826C474A CRC64;

Alignment Scores:  
Pred. No.: 0 Length: 976  
Score: 4451.00 Matches: 845  
Percent Similarity: 91.50% Conservative: 49  
Best Local Similarity: 86.49% Mismatches: 73  
Query Match: 65.72% Indels: 10  
DB: 4 Gaps: 6

US-09-302-812-5 (1-3814) x Q72742 (1-976)

QY 28 ATGAGTGGCGGCGCGCTGGAGCCCTGCACGAAA---GGCGCTGGGCGCGCTGGA 84  
DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTrpGlyAlaAla--- 19  
QY 85 ACTTCTCGCGAGTGGCTCGGACTCCCGAGCTTCCCTGGCAGGACAGCGCGTGTCTC 144  
DB 20 ThrThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnA-ArgValLeu 39  
QY 145 GACCCCAAGACGCTCCCGTCCAGTTCAGGCTCCCTCGCTCCGCGAGCTCGCTCG 204  
DB 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59  
QY 205 GGGCGGCGGAGCCGACAGAGGCAACGCCACCTCGTGTGTTTTCACAAACAAAGACTATT 264  
DB 60 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
QY 265 ACTACTTGGATGATCTAAGGACCCCAAGACACCTGAATCAGAA-----AGTAAA 315  
DB 80 ThrSerTrpMetAspThrLysGlyLysThrAlaGluSerGluSerLeuAspSerLys 99  
QY 316 GAAACCAACATACAGAATTCATCCATGATGAGTTCGTGCAGAAAGATACTTTTAC 375  
DB 100 GluAsnAsnThrArgLleGluSerMetSerSerValGlnLysAspAsnPheTr 119  
QY 376 CCACATAGGTGGAAAAATTGGAAAAATGCTCCTAGCTAAATCTGATTAATACCCACA 435  
DB 120 GlnHisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThr 139  
QY 436 GAAAGAGTTCACAGTATTTGACCAACACAGCAGACTGGAGTGTGCAAGTGGCAGAA 495  
DB 140 GluLysSerThrGlnTy-LeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
QY 496 GAAGCAAGCATGCAGAACAGCTTTTGGCAAGTGGAGCTCCCGCGGAGCTCCGCTACCA 555  
DB 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
QY 556 AAGCAGCTTAGTAATGCTAAACATTGGTCAGTCACCCACACTGATGACCCAGTGACACA 615  
DB 180 GluGlnPheSerAsnAlaAsnLysAspArgSerProGlnAsnAspHisSerAspThr 199  
QY 616 GATCATGAGAGACAGACAGCAATCAGCAGTTCTTACACCTATATAAACTTGCAATACA 675  
DB 200 AspSerGluGluAsnArgAspAsnGlnGlnPheLeuThrValLysLeuAlaAsnAla 219  
QY 676 AAGCCCAACAGTAGGAGATGGGAGCCCGA-----AGCAACTGTAAGTGGAGTGA 726  
DB 220 LysGlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 727 TCTCGCGAGCTGTGAAAGACTGTACAGCTGTCAACAGGAGGAGTGTGATGCTGTACCA 786  
DB 240 SerCysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluLleAspValValPro 259  
QY 787 GAGAGTCTTTCTCAGATCTGTGTCGCGAGGACATTGGAACTGGACCAAAAAATGACAC 846  
DB 260 GluSerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsn 279  
QY 847 AAATTCACCTGGACAGAAAGACGCTTAGGTGATTCGCTCCATTTTGAGAAAGAAAGTGAG 906  
DB 280 LysLeuThrArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299

QY 907 CCTGAGTCCACCAATGGATGTAGACAACTCGAGAAACAGATTGTCAAGATTGAGAGCAGAT 966  
DB 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 967 GAAGAAACAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACACAGCAAT 1023  
DB 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
QY 1024 AAACTTTCAAGTTGCCAAGCAAGAAAGCTGTAGGCGATCTTAGGAAACGGTATTGACT 1083  
DB 340 LysProSerArgPheGlnAlaArgAspAlaAspLleGluPheArgLysArgTrpSerThr 359  
QY 1084 AAGGGAAGTGAAGTTGATTTCCATTTCCAAATTTGAA---GGAGAAATATATGCTGGACC 1140  
DB 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1141 AGTGACTTAAATGCCAAGCCATCTGAAAACCTCTTCTAGCCTTAATGTAGAGTGTAGAAGT 1200  
DB 380 AsnAspLeuAsnAlaLysLeuProGlyAsnLysSerSerLeuAsnValGluCysArgAsn 399  
QY 1201 TCCAAAGCAGCATCGAAAAAGGATTTCTAAATTTACAGATCATTTTCATGAAATTTCCAAG 1260  
DB 400 SerLysGlnHisGlyLysLysAspSerLysLleThrAspHisPheMetArgLeuProLys 419  
QY 1261 TCAGAGGACAGAAAGAAACAATGTGAAGTCAGCATCAAAAGACAGAAAGAGAGATT 1320  
DB 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1321 CCAAAATACATCCCACTAACCTCCCTCCAGAGAAAGAGTGGGAACTCTTATGAG 1380  
DB 440 ProLysTrpValProHisLeuSerProAspLysLysTrpLeuGlyThrProLleGlu 459  
QY 1381 GAAATGAGAAAAATCGCTCGTGGGATCCATTTGCTTTCCTTAAGACCATCTGCAAGT 1440  
DB 460 GluMetArgArgMetProArgCysGlyLleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1441 CACACAGTGTCTGTGGGTAGACCTTCTCAGAGCAGAGAGAGTTCGGAACCTTTTCCA 1500  
DB 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1501 ACATTTACAAAGATTTGTGGGATAACAACATCTGAAATGCTTGTTCGGAACAAAC 1560  
DB 500 ThrHisTrpLysAspLeuTrpAspAsnLysHisValLysMetProCysSerSerGluGlnAsn 519  
QY 1561 TTGTACCTCTGGGAAGATGAGATGTGTGAGCAACTCGAGCAAGTGTGGAGAGTCAAT 1620  
DB 520 LeuTrpValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1621 CAGACTGCCTCTCAACAAATTCACAGCCACAGAACTTGAAGGATCGCATTCGAAA 1680  
DB 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLleLeuLys 559  
QY 1681 TACATCTGGCATATCTTGAAGATGGACTTTACAGCTTTGGTTGTTGTTCTGGGATAG 1740  
DB 560 TyrAsnValAlaTrpSerLysLysTrpAspPheThrAlaLleLleAspPheTrpAspLys 579  
QY 1741 GTACTTGAAGAGCAGAGCGCCCAACATTTATATCAGTCCATTTTACCTGACATGTTGAAA 1800  
DB 580 ValLeuGluGluAlaGluAlaGlnHisLeuTrpGlnSerLleLeuProAspMetValLys 599  
QY 1801 ATTGACTCTGTCTGCCAAATATTGTCACCCAGCCATACCTCCCTGAAACAGAGATG 1860  
DB 600 IleAlaLeuCysLeuProAsnLleCysThrGlnProLleProLeuLeuLysGlnLysMet 619  
QY 1861 AATCATTTCTGTCAGATGTACAGAAACAGATCGCAGTCTTTTAGCTTAATGCTTTCTTC 1920  
DB 620 AsnHisSerIleThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 1921 TGCATATTTCCCGCAGCAATGCCAAGATGAATCGAGTATTCTAGTATTCCAGACATT 1980  
DB 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTrpSerSerTrpProAspIle 659  
QY 1981 AACTTCAATCGGTTGTTTGAAGGAGCTTCAATCAAGAAACAGAAACACTGAAAAACACTC 2040





QY	676	AAGCAACAGTACAGAGATGGCGAGCCAGAG-----AGCAACTGTAGTCCAGTGA	726	QY	1741	GTACTTGAAGAAAGCAGAGAGCCCAACATTTATATCATCTTATCATCTTACCTGACATGGTGAA	1800
Db	220	LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys	239	Db	580	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys	599
QY	727	TCTCGCAGTCTGTGAAGACTGTACAGGCTGTCAACAGGAGGAGGTGGATGTCTACCA	786	QY	1801	ATTGCATCTGTCTGCCAATATTTCCACCAGCCCAATACCATCTCTGAAACAGAGATG	1860
Db	240	SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro	259	Db	600	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	619
QY	787	GAGAGTCTTGTGCAGATGTTGGTGGCAGAGACATTCGAACTGCACCAAAAAATGCAAC	846	QY	1861	AATCATCTGTCTACAGATGTCCAGGAAACAGATCGCCAGTCTTTTAGCTAATCTTCTTC	1920
Db	260	LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn	279	Db	620	AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	639
QY	847	AAATGTGCTGCACAAAGAAAGCAGCTAGGTGATTCGCTCATTTGAGAAAGAGTGA	906	QY	1921	TGCACATTTCCCCCGACGGAATGCCAAGATGAATCGAGTATTCCTAGTACCCAGACATT	1980
Db	280	LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu	299	Db	640	CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle	659
QY	907	CCTGAGTCCCAATGGATGTAGACAACTCGAGAAACAGATGCTCAAGATTTCAGAGCAGAT	966	QY	1981	AATCTCATCTCGTGTGTTGAAGCAGCTTCATCAAGGAAACACAGAAACATGAAAAACATC	2040
Db	300	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	319	Db	660	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu	679
QY	967	GAAGAAACAAGTCCAGTCTTTGATGAGCAAGATGAT---CGTTCTCCCAACAGCAAT	1023	QY	2041	TTCTGCTACTTTCGAGAGTCCAGAGAAACCTACAGATTTGGTGACATTTACAAGA	2100
Db	320	GluGlnThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn	339	Db	680	PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg	699
QY	1024	AAACTTTCAGTGTGCCAAGCAAGAGAGCTGATGCGGATCTTAGAAACGGTATTTGACT	1083	QY	2101	CAGAGTCTTGAAGATTTTCCAGATGCGAAGGTGTGAAAGCCTCTGACACGCTTACAC	2160
Db	340	LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr	359	Db	700	GlnSerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHis	719
QY	1084	AAGGAAGTGAAGTTAGATTGCAATTCCAATTTGAA---GGAGAAATAATGCTGGGACC	1140	QY	2161	GTCACTTACGAGGTACCATAGAAGCAACGCGCAGGATCTACAGGTGGATTTTGA	2220
Db	360	LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet	379	Db	720	ValThrTyrGluGlyThrIleGluGlnGlnGlyGlnGlyMetLeuGlnValAspPheAla	739
QY	1141	AGTCACTTAATGCCAAGCCTCTCGAACTCTTAGCTTAATCTAGCTTAATCTAGAGTGAAGT	1200	QY	2221	AATCGTTTGTGTGAGGTGTGTGACTGTGTCGGGACTTGTACAGAGTATCTGCTGAAACTTAT	2400
Db	380	AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn	399	Db	740	AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe	759
QY	1201	TCCAGAGCAGTGAAGAAAGAGGATCTAAATTAAGATCATTCATGAGATTTCCAG	1260	QY	2281	TTAATCAATCTCGAATTTGATTGTTTTCAGCGCTGTCTACGTGAGTCTGATCACAATCAG	2340
Db	400	SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys	419	Db	760	LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu	779
QY	1261	TCAGAGGACAGAAAGAAAGAACTGTGAGTCAAGATCAAGATCAAGAAAGGAGAT	1320	QY	2341	TGTCTTATATACAGAGTACTGAACAGTACAGTGAATACACAGGCTATCTGCTGAAACTTAT	2400
Db	420	AlaGluAspArgArgLysGluGlnThrLysHisGlnArgThrGluArgLysIle	439	Db	780	CysLeuIleIleThrGlyThrGluGlnThrSerGluTyrThrGlyTyrAlaGluThrTyr	799
QY	1321	CCAAATACATCCACCTAACTCTCCAGAGAGAGTGGCTGGAACTCCTATTGAG	1380	QY	2401	CCTTGGCGCCGAAGCCATGAAGATGGAGTGGAGTGAAGAGACGATTTGGCAGCGCTGCACG	2460
Db	440	ProLysTyrValProProHisLeuSerProAspLysLysThrProIleGlu	459	Db	800	ArgTyrSerArgSerHisLeuAspGlySerGluArgAspCysGluArgCysThr	819
QY	1381	GAAATCAGAAATGCTCGTGGATCCATTTCCTTCCCTTAAGACCATCTGCAAGT	1440	QY	2461	GAGATCGTTGCCATTGAGCGACTTCATTCAGACCGCTACCTCGATCTTGTGCTGCTGAG	2520
Db	460	GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn	479	Db	820	GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu	839
QY	1441	CACACAGTACTCTTCGGGTAGACCTCTGAGAGCAGGAGAGTTCGGAACCTTTTCCA	1500	QY	2521	AAAGTGAGACGTGAGCTTAACAGATCTTCGCGATTCCTCCGCTCGAGTCTCTTCT	2580
Db	480	HisThrValThrIleArgValAspLeuLeuArgLysIleGluValProLysProPhePro	499	Db	840	LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer	859
QY	1501	ACACATTACAAAGATTTGTGGGATAACAACATGTGAAATGCTTGTTCGGAAACAAAC	1560	QY	2581	GAAAACTTTCTGACGTGCCCGCAACCTGGGCTGTGGTGGCTTTGGGGGTGACGCT	2640
Db	500	ThrHisTyrLysAspLeuTyrAspAsnLysHisValLysMetProCysSerGluGlnAsn	519	Db	860	GluAsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAla	879
QY	1561	TTGTACCTCTGGAAGATGAGATGGTGGAGCACTTCAGAGCAGGAGGTTCGGAACCTTTTCCA	1620	QY	2641	AGATTAAGACCTTAATACAGATCTTCGCGAGCTGCTCGCGCTGAACGCTGACGTGGTTAT	2700
Db	520	LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTyrGluLeuIle	539	Db	880	ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla	899
QY	1621	CAGCTGCACTTCTCAACAATTCACACGAGCCCGAGAACTTGAAGATGCGATTCTGAAA	1680	QY	2701	TTTCCCTTTGGGAGCTTCAGAGTGTGAGAGACATTTACAGCATGCACACTTCTCTTACC	2760
Db	540	GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys	559	Db	900	PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr	919
QY	1681	TACAAATGTCATATTTAAGAAATGGGACTTTTACAGCTTTGTGTGATTTCTGGGATAAG	1740	QY	2761	GAGAGAAAGCTGGATGTTGGAAAAGTGTACAGTTHATTGCTTAGTACTACATGAGAA	2820
Db	560	TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuIleAspPheTyrAspLys	579	Db	920	GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu	939
QY	579			QY	2821	TGCAGAAACTGTCTCCACCTCGACAGACATCAAGCTTTATCCATTCATATACCATGCT	2880





[illegible]

88 ProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThrIleAsp 107  
 1603 AGTAGGTGGAGCTCATTCACACTGCACCTCTCCACAAATTCACAGCACCCCAACACTTG 1662  
 108 PheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGluGluLeu 127  
 1663 AAGGATCGGATCTTGAAATACAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTG 1722  
 128 GlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArgAlaLeu 147  
 1723 GNTGATTTCTGGATAAGGTACTTGAGACGACGAGGCCCAACATTTATATCAGTCCATT 1782  
 148 HisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGluAspLeu 167  
 1783 TTACTCTGACATGGTCAAAATTCACCTCTGTCTGCCAAATATTTCACCCAGCCAAATACCA 1842  
 168 LeuProArgGileIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerProValPro 187  
 1843 CTCCTGAAACACAGAAGATGAATCATCTGTCTGCAGATGTCTACAGGACAGACAGATGCCAGTCTT 1902  
 188 LeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSerCysLeu 207  
 1903 TTAGCTAATGCTTTCTTCGCACTTCCCATTTCCCGACGGAATGCC---AAGATGAAATCGGAG 1959  
 208 LeuAlaAsnAlaPheLeuCysThrPheProArgAsnThrLeuLysArgLysSerGlu 227  
 1960 TATCTTAGTTACCCACACATTAACCTCAATCGGTGTTTGAAGGACGTTTCATCAGGAAA 2019  
 228 TyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyProAlaVal 247  
 2020 CCAGAAAACTGAAAAACACTCTCTGCTACTTTTCGAGAGTC-----ACAGAGAAAAA 2073  
 248 LeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGluArgAsp 267  
 2074 -----CCTACAGATTGGTGACATTTACAGACAGAGT-----CTTGAAGAT 2115  
 268 AlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuProGluHis 287  
 2116 TTCCCAAGATGGGAAAGGTGTGAAAGCCTCTG-----ACACGTTACAGCTCACATTAC 2169  
 288 LeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisValAspAla 307  
 2170 GAGGTACCATAGAGGCACGCGCGAGGCATGCTACAGGTGGATTTTGCAAAATCGTTTT 2239  
 308 GluGlyThrIleGluAspGluGlyLeuLeuGlnValAspPheAlaAsnLysTyr 327  
 2230 GTTGAGGTGTGTGACTGTGTGGCGGACTTGTACAAGAAATACAGATTTTATCAAT 2289  
 328 LeuGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheValIleCys 347  
 2290 CCGTAATGATTTTACGGCTGTTCACGTCTTCTACGTGCTGGATCACAATGAGTGTCTTATT 2349  
 348 ProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgProPheGluAlaLeuVal 367  
 2350 ATCACAGTACTGAACAGCTACAGTCAATACACAGGCTATGCTGAACATTATCGTTGGGCC 2409  
 368 MetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGluTrpSer 387  
 2410 CGAAGCCATGAAGATGGGAGTGAAGAAGACGATTGGCAGCGCTGCACGGAGATCGTT 2469  
 388 GlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAlaIleVal 407  
 2470 GCGATTGACGCACTTCATCTACAGCGCTACCTCGATCAGTTTGTGCTGAGAGAAATGAGA 2529  
 408 AlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeuMetGlu 427  
 2530 CGTGAGCTTAAACAAGGCTTTACTGCGGATTCCTCGGT-----CCTGGA 2571  
 428 ArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProProGly 447  
 2572 GTTCCTCTGAAATCTTCTGCAAGTGGCCACGGGAAACTGGGGCTGTGTGCTGCTTGGG 2631

QY 2809 TACAATGAAGATGTCAGAAACTGTTCACCCCTGGACCA-----GAC 2850  
 |||::: |||::: |||:::  
 Db 563 TyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSerLys 582  
 2851 ATCAAGCTTTTCCATTTCATATACCATTCTTGAGTCAAGTCGAGACACCATGCATG 2910  
 |||::: |||::: |||:::  
 QY 583 LysLysLeuTyrAspPheIle-----LysGluGluLeuLysLysValargaspVal 599  
 |||::: |||::: |||:::  
 Db 2911 CCAGGCAGAACGCCA 2925  
 |||::: |||::: |||:::  
 QY 600 ProGlyGluGlyAla 604  
 |||::: |||::: |||:::  
 RESULT 10  
 Q960N8 PRELIMINARY; PRT; 723 AA.  
 ID Q960N8 AC Q960N8; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE LD42380p.  
 GN PARC OR EG:114E2.1 OR CG2864.  
 OS Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 EN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MtKelley;  
 RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celnikier S.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY051955; AAK93379.1; --  
 DR FlyBase; FBgn0023216; PARC.  
 DR InterPro; IPR007724; PARC.  
 DR Pfam; PF05028; PARC; 1.  
 SQ SEQUENCE 723 AA; 81109 MW; 90ASBA59775D611C CRC64;  
  
 Alignment Scores:  
 Pred. No.: 3,79e-81 Length: 723  
 Score: 1069.50 Matches: 243  
 Percent Similarity: 58.05% Conservative: 96  
 Best Local Similarity: 41.61% Mismatches: 194  
 Query Match: 15.79% Indels: 51  
 DB: Gaps: 15  
  
 US-09-302-812-5 (1-3814) x Q960N8 (1-723)  
 QY 1252 ATTTCGAAGTCAGGACAGAGAAGAAACAATGTGAAGTCAGACATCAAAGACAGAA 1311  
 ::::: |||::: |||:::  
 Db 1 MetSerLysSerProaspGlyGlyIleSerGluIleGluThrGluGluGlu----- 17  
 1312 AGGAGATTCCTCAAAAATACATCCCACTTAACCTCCCTCCA-----GAGAAGAAGTGGCTG 1365  
 |||::: |||::: |||:::  
 QY 18 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrpArg 30  
 Db 1366 GGAACTCCTATTGAGGAATGAGAAAAATGCTCGGTGGGATCCATTGGCTTCCTTTA 1425  
 |||::: |||::: |||:::  
 QY 31 GlyValSerMetGluAlaileHisArgValMetTyrGlnLeuProIleArgGluThr 49  
 |||::: |||::: |||:::  
 Db 1426 AGACCA---CTGCGAAGTCACACAGTACTGTTTCGGGTAGACCTCTCGAGACGAGAG 1482  
 |||::: |||::: |||:::  
 QY 50 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGlnLeuProIleArgGluThr 69  
 |||::: |||::: |||:::  
 Db 1483 GTTCCGAAACCTTTTCCACACATTCACAAAGATTTCTGGGATACAAACATGTGAAATG 1542  
 |||::: |||::: |||:::  
 QY 70 ProProArgProTyrLysSerProGlyLys-----TipAspSerGluHisValargLeu 87  
 |||::: |||::: |||:::  
 Db 1543 CTTGTGTCGGAACAAACTTGTCCTCTGGAAGATGAGAATGCTGAGCGAACTGCAGGG 1602  
 |||::: |||::: |||:::

[illegible]



57	Db	AsnMetArgLysSerLeuSerLeuAsnAlaSerAspLeuAla	-----70
1697	QY	CTAAGAATGGGACTTTACAGCTTGGTTCGATTTCTGGATAAGTACTTGAAGAAGCAG	1756
71	Db	ProTyrAlaLeuGlnGlyTyrGly-LeuPheAspLysIleSerArgGluG	89
1757	QY	AGGCCCAACATTTATCAGTCCATTTTACCTGACATGGTGAAATTCGACTCTCTCTGC	1816
89	Db	IuSerAlaAsnPhePheGlyValValProAlaLeuCysArgLeuLeuLeuGlnLeuP	109
1817	QY	CAAAATATTGGCCACGCAATACCACTCTGAAACAGAGATGAATCAT	1866
109	Db	roSerMetLeuGlnLysHisTyr-----GlnLysAlaAspHisValLeuAspG	125
1867	QY	-----TCTGTCAAGATGTCAC	1882
125	Db	lyValLysSerGlyLeuArgLeuLeuGlyProGlnGluAlaGlyIleValLeuLeuSerG	145
1883	QY	AGGAACAGATCCGACAGTCTTTTAGCTAATAGCTTTCTTCGCACATTTCCCGCAGCGAATG	1942
145	Db	InGluLeuIleAlaLeuLeuAlaCysSerPhePheCysLeuPheProGluValAspA	165
1943	QY	CCAAGATGAATCGGAGTATCTAGTATCCAGACATTAACCTCACTCGTCTTTTGA	2001
165	Db	rgSerLeuLys-----AsnLeuGlnGlyIleAsnPheSerGlyLeuPheSerP	181
2002	QY	-----GGACGTTTCATCAAGAAACACAGAAACCTGAAACACCTCTTCGTCTACT	2050
181	Db	heProTyrMetArgHisCysThrLysGlnGluAsnLysIleCysLeuIleHisTyrP	201
2051	QY	TTCCGAAGATGCACAGAGAAACCTCAGGATGGTGACATTTACAAGACAGAGTCTTG	2110
201	Db	heGlyArgIleCysArgTyrMetProThrGlyPheValSerPheGluArgLysIleLeuP	221
2111	QY	AA-----GATTTCCAGAA-----TGGGAAAGGTGTG	2137
221	Db	roLeuGlnTyrHisProHisPheValSerTyrProLysAlaAspSerTyrAlaAsnSerV	241
2138	QY	AAAGCCTCTGACAGCTTACACGTCTACTACGAGGTACATGAGAGCGCAACGCCGAG	2197
241	Db	alThrProLeuCysSerIleGluIleHisThrSerGlyAlaIleGluAspGlnProCysG	261
2198	QY	GCATGCTACAGGTGGATTTTGCAAATCGTTTGTGGAGGTGGTGTGACTGTCGGGAC	2257
261	Db	luAlaLeuGluValAspPheAlaAspGlnTyrPheGlyLeuThrLeuSerTyrAspT	281
2258	QY	TTGTACAGAAGAAATCGATTTTATCAATCTGTAATGATTTTCACGCTGTCTCA	2317
281	Db	hrLeuGlnGlnGluIleArgPheValIleAsnProGluLeuIleAlaGlyMetIlePheL	301
2318	QY	CTGAGTCTCGATCACAAATCAGTCTCTTATTCACAGGTACTGAACAGTACAGTGAAT	2377
301	Db	euProArgMetAspAlaAsnGluAlaIleGluIleValGlyValGluArgPheSerGlyT	321
2378	QY	ACACAGGCTATGCTGAAACTTATCGTTGGGCCGAGCATCAAGATGGAGTGAAGAG	2437
321	Db	YrThrGlyTyrProSerPheGlnTyrAlaGlyAspTyrThrAspAsnLysAspLeuA	341
2438	QY	ACGATTGGCAGCGCGCTGCAGGAGATCGTTGCCATTCACGCATCTTCAGAGCT	2497
341	Db	spIlePheArgArgArgLysThrArgValIleAlaIleAspAlaMetPro-----AspProG	360
2498	QY	ACCTCGATCAGTTTGTCCCTGAGAAGTGAAGCTGAGCTTACAGGCTTACTCGGAT	2557
360	Db	lyMetGlyGlnTyrLysLeuAspAlaLeuIleArgGluValAsnLysAlaPheSerGlyT	380
2558	QY	TCTCCGT-----CCTG	2569
380	Db	yrMethisGlnCysLysTyrAsnIleAspValLysHisAspProGluAlaSerSerSerH	400
2570	QY	GAGTTCTCTGAAATCTTTCTGCA-----	2595
400	Db	isValProLeuThrSerAspSerAlaSerGlnValIleGluSerSerHisArgTrpCysI	420



QY	2596	-----GTGGCCACGGGAACTGGGCTGTGGTGGCTTTG 2629	Query Match:	6.76%	Indels:	98
DB	420	leAspHisGluGluLysLysIleGlyValAlaThrGlyAsnTrpGlyCysGlyValPheG 440	DB:	5	Gaps:	18
QY	2630	GGGGTGACGCTAGATTAAAGCCTTAATACAGATCTGGCAGCTGGCGCTGAACGTG 2689	US-09-302-812-5 (1-3814) x Q9N5L4 (1-485)			
DB	440	lyGlyAspProGluLeuLysIleMetLeuGlnTrpLeuAlaIleSerGlnSerGlyArgP 460	QY	1609	TGGGAGCTCATTCACAGCTGCCTCTCAACAAATTCACACGA-----CCCAGAACCTTGAAG 1665	
QY	2690	AGCTGGTT---TATTTCACCTTTGGGACTCAGATTGATGAGACATTTACAGCATGC 2746	DB	46	PheGluLeuLeuGluThrGlyValSerGlnGlnTrpArgAsnCysAspGlnAenLeuPhe 65	
DB	460	roPheMetSerTyrtyrThrPheGly---LeuGlnAlaLeuGlnAenLeuAenGlnValI 479	QY	1666	GATGGATTCGAAA-----TACAAATGGGCATATTCT- 1698	
QY	2747	ACACTTTCTTACCGAGAGAGCTGATGTTCGAAAAGTGTAAGTTATTGCTTAGAT 2806	DB	66	AsnGluTyLeuLysThrTyrlsAsnGlyGlyTySerGlnPheGluAspLeuLeuPhe 85	
DB	479	leGluMetValAlaLeuGlnGluMetThrValGlyAspLeuTrpLysLysLeuValGluT 499	QY	1699	-----AAGAAATGGCATTTACACGTTTGGTGGTATTTC 1731	
QY	2807	ACTACAATGAA 2817	DB	86	LysIleTrpGlyTySerGluGluLysGluArgPheAspLeuProAlaLeuLysSerPhe 105	
DB	499	yrSerSerGlu 502	QY	1732	TGGGATAAGGTACTTGAA-----GAAGCAGAGGCCCAACATTTATATCATGTCATT 1782	
RESULT 13			DB	106	TyrArgLysMetSerGluIleValGlyGluAspGlu-----Val 118	
ID	Q9N5L4	PRELIMINARY; PRT; 485 AA.	QY	1783	TTACTTGACATGGTGAATAATTCACCTCTCTGCCAAATATTTGCCACCCAGCCAATACCA 1842	
AC	Q9N5L4;		DB	119	LeuGluLysLeuAlaArgLeuValArgIleThrLysSerAlaCys-----Glu 134	
DT	01-OCT-2000	(TrEMBLrel. 15, Created)	QY	1843	CTCCTGAAACAGAGATGAAT-----CATCTGTCAAGCATGTCACAG 1884	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	DB	135	ValLeuProGluLysIleTyrArgLeuValGlyAspIleGluSerAlaThrPheSerHis 154	
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)	QY	1885	GAACATATCGCCAGTCTTTTGTAGTAATGCTTTCTTGTGCACATTTCCCGACGGAATGCC 1944	
DE	Hypothetical protein H23L24.5 (Poly ADP-ribose metabolism enzyme 4).		DB	155	IleGlnCysAlaSerLeuIleAlaTrpMetPhePheSerAspThrProArg----- 171	
GN	H23L24.5 OR PME-4.		QY	1945	ARGATGAATCGGAGTATTCTAGTTACCCAGACATTAACCTCAATCGTGTGTTGAAGGA 2004	
OS	Caenorhabditis elegans.		DB	172	-----LeuSerPheIleIleLeuGlnLys 180	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		QY	2005	CGTTCATCAAGGAACCCAGAAACTGTAACACTCTTCTGTACTTTCGAAGAGTGCACA 2064	
OC	Rhabditidae; Peloderidae; Caenorhabditis.		DB	181	ThrThrCysValAlaValGluLysLeuLysPheLeuPheThrThrPheAspLysMetSer 200	
OX	NCBI_TaxID=6239;		QY	2065	GAGAAAAAACCCTACAGGATTCGTGACATTTTACA-----AGACAGAGT 2106	
RN	SEQUENCE FROM N.A.		DB	201	IleAspProIleGlyAlaValSerPheArgLysMetArgIleThrHisGlnTrp 220	
RP	STRAIN=Bristol N2;		QY	2107	CTTGAAATTTCCAGAAATGGGA-----AGTGTGAAAAGCCTCTGACACGCTTACACGTC 2163	
RC	STRAIN=Bristol N2;		DB	221	LeuGluAen-----TrpLysLeuArgLysThrAsnLeuLeuProAspValGlnVal 237	
RA	Abu-Threideh J., Lehnert L.;		QY	2164	ACTTACGAGGTACCATAGAGGCAACGGCGAGGCATGCTACAGGTGGATTTTGCAAT 2223	
RT	"The sequence of C. elegans cosmid H23L24.";		DB	238	PheAspLysMetSerIleGluGluThrAlaLeu-----CysThrGlnIleAspPheAlaAen 256	
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.		QY	2224	CGTTTGTGGAGGTGGTGTGACTGGTGGCGGACTTGTACAGAGAAATCAGATTTTGA 2283	
EN	[2]		DB	257	LysArgLeuGlyGlyGlyValLeuLysGlyAlaValGlnGluGluIleArgPheMet 276	
RP	SEQUENCE FROM N.A.		QY	2284	ATCAATCTGAATTTGTTTTCACGCTGTTCACGCTGTTCACGAGGTGGTGGATCACAATGAGTGT 2343	
RC	STRAIN=Bristol N2;		DB	277	MetCysProGluMetMetValAlaIleLeuLeuAenAspValThrGlnAspLeuGluAla 296	
RA	Waterston R.;		QY	2344	CTTATTATCAGAGTACTGACACATGATGAAATACACAGGCTATGCTGAAACTTATCGT 2403	
RT	Investigating biology. The C. elegans Sequencing Consortium.";		DB	297	IleSerIleValGlyAlaTyrValPheSerSerTyrlsGlyTySerAsnThrLeuLys 316	
RL	Science 282:2012-2018(1998).		QY	2404	TGGGCCCGA-----AGCCATGAAGATGGAGTGAAGACGATGG 2445	
EN	[3]		DB	317	TrpAlaLysIleThrProLysHisSerAlaGlnAenAenSerPheArgAspGlnPhe 336	
RP	SEQUENCE FROM N.A.		QY	2446	CAGCGCGCTGCAGGAGATCGTTGCCATTCAGGCACCTTCAC-----TTC 2490	
RC	STRAIN=Bristol N2;		DB	337	GlyArgLeuGlnThrGluThrValAlaIleAspAlaValArgAsnAlaGlyThrProLeu 356	
RA	Waterston R.;					
RT	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.					
RL	[4]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=Bristol N2;					
RA	Gagnon S.N., Hardy I., Desnoyers S.;					
RT	"Molecular characterization of poly ADP-ribose glycohydrolases in Caenorhabditis elegans.";					
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AC006662; AAF39896.2; -					
DR	EMBL; AF548468; AAN40699.1; -					
DR	WormPep; H23L24.5; CE32685.					
DR	InterPro; IPR007724; PARG.					
DR	Pfam; PF05028; PARG; 1.					
KW	Hypothetical protein.					
SQ	SEQUENCE 485 AA; 55574 MW; 86367DDBAB15081F CRC64;					
Alignment Scores:						
Pred. No.:	1,46e-29	Length:	485			
Score:	458.00	Matches:	142			
Percent Similarity:	46.69%	Conservative:	84			
Best Local Similarity:	29.34%	Mismatches:	160			

Thu May 27 09:56:11 2004

Percent Similarity: 38.79% Conservative: 137  
 Best Local Similarity: 22.88% Mismatches: 297  
 Query Match: 6.66% Indels: 230  
 DB: 5 Gaps: 37

US-09-302-812-5 (1-3814) x Q867X0 (1-781)

QY 622 GAAGAGACAGACAGCAATCAGCAGTTTCTTACACCTATATAAACTTCAAAATCAAAAGCCA 681  
 DB 14 GlnAspGluLeuSerPheLeuVal---GlyValGlyPheAlaHisGlnValPro 32  
 QY 682 ACAGTAGCAGATGGCAGCCAGCAAGCAAACTGTAAGTCAGTCGATCGCCAGTCGTG 741  
 DB 33 ThrMet-----LysArgGlyLeuThrGluHisGlyAsnThrThrGluSerLys 49  
 QY 742 AAAGACTGTACGGCTGTCAACAGAGAGAGGTGTCTACACAGAGATCTCTTGTCTCA 801  
 DB 50 GluAspProGlu-----GluProLysSerArgAspValPheValSerSerGlnSer 67  
 QY 802 GATGTTGGTCCGAGGACATTCGAACTGGACCAAAA----- 837  
 DB 68 AspGluSerGlnGluAspSerAlaGluAsnProGluLeuAlaLysGluValSerGluAsn 87  
 QY 838 -----AATGACAAACAAATTG 852  
 DB 88 CysGluAsnLeuThrGluThrLeuLysIleSerAsnIleGluSerLeuAspAsnValThr 107  
 QY 853 ACTGGACAAAGACAGCCCTAGGTGAT-----TCGCTCCATTGTGAGAAAGT 903  
 DB 108 GluArgSerGluHisThrLeuAspAsnHisLysSerThrGluProMetGluGluAspVal 127  
 QY 904 GAGCTGAGTCACCAATGATGTAGACACTCGAGAAACAGTTCTCAAGATTCAGAA--- 960  
 DB 128 AsnAsnLysSerAsnIleAspVal-----AlaIleAsnSerAspGluAspGluLeu 145  
 QY 961 -----GCAGATGAAGAAACAAAGTCCAGTCTTT 987  
 DB 146 ValLeuGluGluAsnAsnLysGluMetArgAspGlyGluGlnValGlnGlnAspLeuPhe 165  
 QY 988 ---GATGACGACAGAT----- 999  
 DB 166 AlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAspThrThrThrGlnLeu 185  
 QY 1000 -----GATCGTTCTCCCAACAGCAATAAATTTCAAGTTGCAAGCAGAGAA 1050  
 DB 186 AspIleThrAspSerGluValGluThrAlaGlnLysMetGluMetIleGluGluThrGlu 205  
 QY 1051 GCTGATGGCGATCTTAGAAACGGTATTTGACTAAGGAAGTGAAGTTAGATTGCAATTC 1110  
 DB 206 AlaAspSer----- 208  
 QY 1111 CAATTTGAAGGAGAAATAATCTGGACCACTGACTTAATATGCCAAGCATCTGGAAC 1170  
 DB 209 ThrPheValGlyGluAspSerLys-----AsnGlnArgGlnSerGlyThr 223  
 QY 1171 TCT-----TCTACCTTATGTAGAGTGTAGAGTTCCTCAAGCAG 1209  
 DB 224 ThrSerAspGluValAspAlaAspSerGlnIleAsnLeuAlaThrLysThrValArg--- 242  
 QY 1210 CATGGAAGAGGATCTAAATTAATCAGATCATTTTCATGAGATTTTCCAGTCAGAGGAC 1269  
 DB 243 -----ThrSerSerSerPheLeuSerThrValSerThrCysGluAla 257  
 QY 1270 AGAAGAAAGAACCAATGTGAAGTCAGACATCAAGAACAGAAAGGAGATTCCAAATAC 1329  
 DB 258 ProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPhe 277  
 QY 1330 ATCCACCTAACCTCCCTCCAGAGAGAGAGTGGCTGGGACTCTCTATTGAGGAATACGA 1389  
 DB 278 ThrGluGlyAsnLeu----- 282  
 QY 1390 AAATGCTCGGTGGTGGATTCCTTTCCTTAAGACCATCTGCAAGTCTCACAGTG 1449

2491 AGACGCTACTCGATCAGTTTGTGCTGAGAAAGTACAGAGCTTAACAGGCTTAC 2550  
 DB 357 GluCysLeuLeuAsnGlnLeuThrThrGluLysLeuThrArgGluValArgLysAlaAla 376  
 QY 2551 TGCGGATTCCTCGTCTCTGGA-----GTTCTTCTGAAATCTTTCTGCA 2595  
 DB 377 IleGlyPheLeuSerAlaGlyAspGlyPheSerLysIlePro----- 390  
 QY 2596 GTGGCCACAGGAAACCTGGGGCTGTGTGCTTTGGGGGTGACGCTAGATTAAAGCCTTA 2655  
 DB 391 ValValSerGlyTrpGlyCysGlyAlaPheArgGlyAsnLysProLeuLysPheLeu 410  
 QY 2656 ATACAGATCTCGACAGCTGTGCGGCTGACGTGACGTGCTTTATTTTCACCTTTGGGAC 2715  
 DB 411 IleGlnValIleAlaCysGlyIleSerAspArgProLeuGlnPheCysThrPheGlyAsp 430  
 QY 2716 TCAGAGTTGATGAGACAGACATTTACAGCATGCACACTTTCTCCGAGAGAGAGCTGGAT 2775  
 DB 431 ThrGluLeuAlaLysLysCysGluGluMetMetThrLeuPheArgAsnAsnValArg 450  
 QY 2776 GTTGAAAAGTGTACAGATTATTCCTTAGATACACTACATGAAGATGCAGAACTGTTC 2835  
 DB 451 ThrGlyGlnLeuPheLeuIle-----Asn 459  
 QY 2836 ACCCTCGACAGACATCAAGCTT---TATCCATTATATACCAAGCTGTGAGTCAAGT 2892  
 DB 460 SerIleGlyProProLeuAsnTyrSerGluGlnTyrValPheAspAlaIleArgAlaLys 479  
 QY 2893 GCAGAGACCACT 2904  
 DB 480 IleAsnSerThr 483

RESULT 14  
 Q867X0 PRELIMINARY; PRT; 781 AA.  
 AC Q867X0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Poly ADP-ribose metabolism enzyme-3 long form (C. elegans PME-3  
 protein) (Corresponding sequence F20C5.1a).  
 CN F20C5.1 OR PME-3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gagnon S.N., Hardy I., Desnoyers S.;  
 RT "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
 Caenorhabditis elegans.";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 investigating biology.";  
 RL Science 282:2012-2018 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Matthews P.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY185493; AA026316.1; --  
 DR EMBL; Z68161; CAD89735.1; --  
 DR WormPep; F20C5.1a; CE33775.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 SQ SEQUENCE 781 AA; 89273 MW; 30447D06605CED0A CRC64;

Alignment Scores: 6.81e-29 Length: 781  
 Pred. No.: 197  
 Score: 451.00

283 Db -----ThrLeuGlnProAspLeuAen----- 289  
 1450 QY ACTGTTCCGGTAGAC-----CTTCTCAGCAGCAGGAGGTTTCGAAACCTTTTCCA 1500  
 290 Db -----LysValAspProAspArgAsnTyrArgTyrCysThrIleProAen---PhePro 306  
 1501 QY ACACATTACAAGATTG---TGCGATAACAAACAT-----GTGAAATGCTTGT 1548  
 307 Db AlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLysIleValLeuPro--- 325  
 1549 QY TCGGAACAAAACCTTGTACCTCTGTGGAAGATGAGAACTGCGAGGAGTAGG 1608  
 326 Db -----GlnArg 327  
 1609 QY TGGAGCTCATTCAGACTGCCTTCTCAACAAATTCACACGACCCGAACTTGAAGCAT 1668  
 328 Db Trp-----ArgGluPheAspSerArgGlyArgArgAsp 339  
 1669 QY CGGATTCTGAAATACAAAT-----GTGCAATATCTAAGAAATGGGACTTTACAGCT 1719  
 340 Db SerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCysTyrLysThrThrGly 359  
 1720 QY -----TTGTTGATTCTCGGAT-----AAGTACTTGAA 1749  
 360 Db TyrPheMetPheValGlyLeuLeuHisAsnMetTyrGluPheAspProAspIleThrTyr 379  
 1750 QY GAAGCAGAGGCCCAACATTTATATACATTCATTTACCTGACATCGTG----- 1797  
 380 Db LysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeuValGlyArgGluGlu 399  
 1798 QY -----AAATTCACCTCTGCTGCTGCCAAATATTTGCACC 1830  
 400 Db ValLeuGluLysPheAlaArgValAlaAlaArgIleAlaLysThrAlaGluAspIleLeuPro 419  
 1831 QY CAGCCAAATACCTCTGTAACAGAGATGATCATTTCTGTCAGCATGTCACAGAACAG 1890  
 420 Db GluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThrLeuSerHisLysGln 438  
 1891 QY ATCCGAGCTTTTACGTAATGCTTTCTCTGACATCTTCCCGACGGAATGCCAGATG 1950  
 439 Db CysAlaAlaLeuValAlaArgMetPhePhe-----AlaArgPro 451  
 1951 QY AAATCCGAGTATCTAGTTACCCAGACATTAACCTCAATCGGTTGTTGAAGACATTTCA 2010  
 452 Db AspSerProPheSer-----PheCysArgIleLeuSerSerAspLys 465  
 2011 QY TCAAGGAACCCAGAAACCTGAAACACTTTCTGCTACTTTCGAGAGTTCACAGAGAAA 2070  
 466 Db SerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAspLysMetSerMetAsp 485  
 2071 QY AAACCTACAGGATGTGCATTT-----ACAAGACAGAGTCTGAAAGATTTT---CCA 2121  
 486 Db ProProAspGlyAlaValSerPheArgLeuThrLysMetAspLysAspThrPheAsnGlu 505  
 2122 QY GAATGG---GAAAGGTGTGAAGCCCTCTCAGCAGCTTACACGCTTACACGAGGTACC 2178  
 506 Db GluTyrLysAspLysLysLeuArgSerLeuProGluValGluPhePheAspGluMetLeu 525  
 2179 QY ATAGAGGCCAACCGCCGAGGATGCTACAGTGTGATTTTCGAATCGTTTGTGGAGGT 2238  
 526 Db IleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsnGluHisLeuGlyGly 544  
 2239 QY GGTGTGATGTGGGGACTGTACAGAGAAATCAGATTTTAAATCAATCTCTGAATG 2298  
 545 Db GlyValLeuAsnHisGlySerValGlnGluIleArgPheLeuMetCysProGluMet 564  
 2299 QY ATTGTTTTCAGCGCTGTTCTACCTGAGGTGCTGGATCACAATGAGTGTCTTATATCAGAGT 2358  
 565 Db MetValGlyMetLeuLeuLysGluLysMetLysGlnLeuGluAlaIleSerIleValGly 584  
 2359 QY ACTGAACAGTACAGTGAATACAGGCTTATCTGAACTTATCTGCGGCC----- 2409  
 585 Db AlaTyrValPheSerSerTyrTyrGlyThrGlyHisThrLeuLysThrPalaGluLeuGln 604

2410 QY -----CGAAGCCATCAAGATGGGAGTGAA---AAGACGATTTGGCAGCGCGCTGCACG 2460  
 605 Db ProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPheGlyArgLeuArgVal 624  
 2461 QY GAGATCGTTGCTCCATTCAGCAGCTTCACTTCCAGACGC-----TACCTCGAT 2505  
 625 Db GluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGlu 644  
 2506 QY CAGTTTGTGCTGAGAAAGTGAAGCTGAGCTTAAACAAGCTTACTCGGATTCCTCGT 2565  
 645 Db GlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSerIleGlyPheMetSer 664  
 2566 QY CCTGAGTTCCTCTGAAAATCTTCTGCACTGCGCCACGGGAAACTGGGGCTGTGGTGCC 2625  
 665 Db GlnGlyProLysPheThrAsnIle---ProIleValThrGlyTyrTrpGlyCysGlyAla 683  
 2626 QY TTTGGGGTGCAGCTAGATTAAGAGCTTAATACAGATCTCGGAGCTGCTCGGCTGAA 2685  
 684 Db PheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAsp 703  
 2686 QY CGTGAAGTGTGTTTATTTCCCTTTGGGACTCTCAGAGTTGATGAGAGACATTTACAGCATG 2745  
 704 Db ArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAlaLysCysLysLysIle 723  
 2746 QY CACACTTCTCCGAGAGGAAGCTGGATGTTGGAAAAGTGTACAGTTATTGCTTAGA 2805  
 724 Db IleGluArgMetLysGlnLysAspValThrLeuGlyMetLeuPheSerMetIle----- 741  
 2806 QY TACTACAATGAAGATGCAGAAACTGTTCCACCTCGGACCCAGACATCAAGCTTTATCCA 2865  
 742 Db -----AsnAsnThrGlyLeuProHisLysHisPheGluPheTyrVal 755  
 2866 QY TTC 2868  
 756 Db Phe 756  
 RESULT 15  
 Q19637 ID Q19637 PRELIMINARY; PRT; 764 AA.  
 AC Q19637;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE PME-3 protein (corresponding sequence F20C5.1) (poly ADP-ribose  
 metabolism enzyme-3 short form).  
 DE F20C5.1 OR PME-3.  
 GN Caenorhabditis elegans.  
 OS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Matthews P.;  
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA none;  
 RL "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology.";  
 RL Science 282:2012-2018 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Gagnon S.N., Hardy I., Desnovers S.;  
 RL "Characterization of poly(ADP-ribose) glycohydrolases in the nematode  
 Caenorhabditis elegans.";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z68161; CAA92299.2; --  
 DR EMBL; AY185494; AAC26317.1; --  
 DR PIR; T21138; T21138.  
 DR WormPep; F20C5.1b; CE32867.  
 DR InterPro; IPR007724; PARG.





Thu May 27 09:56:11 2004

us-09-302-812-6.rag

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 110.904 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-6  
Perfect score: 5164  
Sequence: 1 MSAGPGEPCYKARWAGT.....YHAVESSAETTDMPQKAGT 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	5164	100.0	968	5	AAE25631	Murine po
2	5164	100.0	968	5	AAU76022	Mouse pol
3	5164	100.0	968	5	AAU76013	Mouse pol
4	5164	100.0	968	6	ABG72280	Mmurine p
5	4395	85.1	976	5	AAE25630	Human pol
6	4395	85.1	976	5	AAU76021	Human pol
7	4395	85.1	976	5	AAU76012	Human pol
8	4395	85.1	976	6	ABG72279	Human pol
9	4332.5	83.9	977	5	AAE25629	Bovine po
10	4332.5	83.9	977	5	AAU76020	Bovine po
11	4332.5	83.9	977	5	AAU75799	Bovine po
12	4332.5	83.9	977	6	ABG72278	Bovine po
13	1074.5	20.8	768	4	ABBS9491	Drosophil
14	1074.5	20.8	768	5	AAE25632	Fruit fly
15	1074.5	20.8	768	5	AAU76023	Fruit fly
16	1074.5	20.8	768	5	AAU76014	Fruit fly
17	1074.5	20.8	768	6	ABG72281	Fruit fly
18	443.5	8.6	726	5	AAE25633	Poly aden
19	443.5	8.6	726	5	AAU76024	Worm poly
20	443.5	8.6	726	5	AAU76015	Worm poly
21	443.5	8.6	726	6	ABG72282	C. elegans
22	350.5	6.8	819	4	ABG20721	Novel hum
23	199	3.9	100	4	ABG11103	Novel hum
24	186	3.6	33	5	AAE25652	Bovine po
25	178.5	3.5	931	4	ABB61093	Drosophil

26	177	3.4	2703	4	ABB60074	Drosophil
27	172	3.3	1489	6	ABO14779	Novel hum
28	172	3.3	1489	6	ABU11038	Human pro
29	169.5	3.3	1297	6	ABU43768	Protein e
30	168	3.3	335	4	ABG09631	Novel hum
31	168	3.3	2951	4	ABG60291	Drosophil
32	167	3.2	31	5	AAE25634	Bovine po
33	167	3.2	31	5	AAE25651	Bovine po
34	167	3.2	31	5	AAU76025	Bovine po
35	167	3.2	31	5	AAU76016	Bovine po
36	167	3.2	31	6	ABG72283	Oligopept
37	167	3.2	1169	6	ABU43765	Protein e
38	167	3.2	1177	5	ABP40261	Staphyloc
39	165.5	3.2	787	6	ABP55414	Human MDD
40	165	3.2	1192	3	ABE18165	Plasmidiu
41	164.5	3.2	805	3	ABP55404	Human MDD
42	163.5	3.2	779	3	AAE42438	Human ORF
43	163.5	3.2	1187	2	AAE66451	AF-4 prot
44	163.5	3.2	1210	2	AAE66450	AF-4 prot
45	163.5	3.2	1210	7	ABR63878	Human AF4

ALIGNMENTS

RESULT 1  
AAE25631  
ID AAE25631 standard; protein; 968 AA.

XX AAE25631;

XX 04-NOV-2002 (first entry)

XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX Murine poly adenosine diphosphate-ribose glycohydrolase; PARG, enzyme;  
ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiant;  
cytostatic; vasotrophic; neuroprotective; anticonvulsant; gene therapy;  
antisenase therapy.

XX Mus musculus.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42083.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 3; Col 63-68; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
CC disease. The invention is useful in gene therapy and antisense therapy.  
XX The present sequence is murine PARG  
SQ

Query Match		100.0%;	Score 5164;	DB 5;	Length 968;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 968;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSAGPWGPECTKARWGAAGTSAPTASDRSPFQRRVLDPKAPVQFRPPSPACVSG	60		
DB	1	MSAGPWGPECTKARWGAAGTSAPTASDRSPFQRRVLDPKAPVQFRPPSPACVSG	60		
QY	61	RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESSEKNNNTRIDSMSSVQKONFYPHKV	120		
DB	61	RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESSEKNNNTRIDSMSSVQKONFYPHKV	120		
QY	121	EKENVPQLNDKSPTEKSSOYLNQOQTASVCKWQNEGKHAQLLASEPPAGTLPKQLS	180		
DB	121	EKENVPQLNDKSPTEKSSOYLNQOQTASVCKWQNEGKHAQLLASEPPAGTLPKQLS	180		
QY	181	NANIGQSPHTDDHSDTHDEDRDNOQFLTPKLANTKPTVGDQARSNCCKSGRSQVKD	240		
DB	181	NANIGQSPHTDDHSDTHDEDRDNOQFLTPKLANTKPTVGDQARSNCCKSGRSQVKD	240		
QY	241	CTGCOQEEVVLPSPLSDVGAEDGTGPKNDKLTGOESSLGDSPPPEKESESESPMDV	300		
DB	241	CTGCOQEEVVLPSPLSDVGAEDGTGPKNDKLTGOESSLGDSPPPEKESESESPMDV	300		
QY	301	DNSRNSCQSEADETSPVDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGKSEVRLH	360		
DB	301	DNSRNSCQSEADETSPVDEQDDRSSQTANKLSSCQAREADGDLRKRYLTGKSEVRLH	360		
QY	361	FOFEGENNAGTSDLNKAPSGNSSLNVECRSKQHGKRDSDKITDHFMRISKSEDRKQEC	420		
DB	361	FOFEGENNAGTSDLNKAPSGNSSLNVECRSKQHGKRDSDKITDHFMRISKSEDRKQEC	420		
QY	421	EVRHQRTERKIPIKYPNLPPEKKMLGTPIEMRWKMPRCGIHLPSLRPSASHTVVRVDL	480		
DB	421	EVRHQRTERKIPIKYPNLPPEKKMLGTPIEMRWKMPRCGIHLPSLRPSASHTVVRVDL	480		
QY	481	LRAGEVPKFPFTHYKDLNDKWKMPSCSQNLVPVEDENGERTAGSRWELIQTALLNKFT	540		
DB	481	LRAGEVPKFPFTHYKDLNDKWKMPSCSQNLVPVEDENGERTAGSRWELIQTALLNKFT	540		
QY	541	RPQNLKADAILKXNVAYSKKWDFTALVDVFDKVLLEAEAGHLVQSILPDMVKIALCLPNLC	600		
DB	541	RPQNLKADAILKXNVAYSKKWDFTALVDVFDKVLLEAEAGHLVQSILPDMVKIALCLPNLC	600		
QY	601	TQIPILLKQKNNHVSVMQSQIOLASLANAFCTFPPRNAMKSEYSYDIPNFRNLFEGR	660		
DB	601	TQIPILLKQKNNHVSVMQSQIOLASLANAFCTFPPRNAMKSEYSYDIPNFRNLFEGR	660		
QY	661	SSRKEPKLTLFCYFRVATKPKTGLVTFTRQSLDFPWECEKPLTRLHVYEGTIEG	720		
DB	661	SSRKEPKLTLFCYFRVATKPKTGLVTFTRQSLDFPWECEKPLTRLHVYEGTIEG	720		
QY	721	NGRGLQVDFANRFGVGGVTGAGLVQBEIRFLNPELIVSRFLTEVLHDNECLLIITGTEQ	780		
DB	721	NGRGLQVDFANRFGVGGVTGAGLVQBEIRFLNPELIVSRFLTEVLHDNECLLIITGTEQ	780		
QY	781	YSEYTGAEYTRWARSHEGSKDDWQRCTEIVAIIDALHFRYLDQFVPEKVRRELNKA	840		
DB	781	YSEYTGAEYTRWARSHEGSKDDWQRCTEIVAIIDALHFRYLDQFVPEKVRRELNKA	840		
QY	841	YCGFLRPGVPSENLSAVATGNWGGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM	900		
DB	841	YCGFLRPGVPSENLSAVATGNWGGCGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM	900		

QY 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECNCSCTGPDIKLYPFIYHAYESSAETTD 960  
DB 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECNCSCTGPDIKLYPFIYHAYESSAETTD 960  
QY 961 MFGQKAGT 968  
DB 961 MFGQKAGT 968

RESULT 2

AAU76022	AAU76022 standard; protein; 968 AA.
XX	AAU76022;
AC	08-MAY-2002 (first entry)
DT	Mouse poly (ADP-ribose) glycohydrolase (PARG) protein sequence.
DE	Mouse; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;
KW	adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;
KW	leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;
KW	hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;
KW	neurodegenerative disease; neurological disorder; Alzheimer's disease;
KW	Huntington's disease; Parkinson's disease.
OS	Mus musculus.
XX	US6337202-B1.
PN	08-JAN-2002.
FD	23-FEB-2000; 2000US-00511477.
PF	01-MAY-1998; 98US-0083768P.
XX	30-APR-1999; 99US-00302812.
PR	(KENT) UNIV KENTUCKY RES FOUND.
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;
PA	WPI; 2002-163240/21.
PI	N-PSDB; ABK14933.
DR	Novel isolated and purified poly (ADP-ribose) glycohydrolase protein
DR	which catalyses release of ADP-ribose from ADP ribose polymer, useful for
XX	treating neoplastic and neurological disorders, heart attack and stroke.
PT	Claim 2; Col 63-70; 81pp; English.
PS	The present invention relates to a new poly (ADP-ribose) glycohydrolase
XX	(PARG) protein which catalyses release of ADP-ribose from an ADP
CC	(adenosine diphosphate)-ribose polymer. The PARG molecule of the
CC	invention is useful for generating antibodies and can be inhibited or
CC	activated for diagnosing and treating neoplastic disorders such as
CC	adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC	teratocarcinoma, hyperplasia and hypertrophy, reperfusion following
CC	ischaemia, heart attack, stroke, neurodegenerative diseases, neurological
CC	disorders including Alzheimer's, Huntington's and Parkinson's diseases,
CC	and related conditions. PARG levels may be enhanced to suppress DNA
CC	repair and increase the cell's susceptibility to chemotherapy drugs.
CC	Antagonists of PARG are administered to treat or prevent neoplastic
CC	disorders. The present amino acid sequence represents the mouse PARG
CC	protein of the invention. This protein is one of several PARG proteins
CC	(AAU76020-AAU76024) of the invention
XX	Sequence 968 AA;
SQ	Query Match 100.0%; Score 5164; DB 5; Length 968;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu May 27 09:56:11 2004

AC XX AC AAU76013;  
 DT XX 08-MAY-2002 (first entry)  
 DE XX Mouse poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX XX Mouse; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX XX Mus musculus.  
 OS US6333148-B1.  
 PN 25-DEC-2001.  
 XX 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson BL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14495.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX Claim 3; Col 63-68; 80pp; English.  
 XX The present invention relates to a new method for screening compounds for  
 XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 XX are used to treat or prevent any condition associated with DNA damage,  
 XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 XX Compounds identified by the new method are more effective than known  
 XX inhibitors and have fewer side effects. The present amino acid sequence  
 XX represents the mouse PARG protein of the invention. This protein is one  
 XX of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 XX invention  
 XX SQ Sequence 968 AA;  
 Query Match 100.0%; Score 5164; DB 5; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSAGPGWECTKARWGAAGTSAPTASDSRSPGRRVLDPKDAPVQFRVPPSPACVSG 60  
 Db 1 MSAGPGWECTKARWGAAGTSAPTASDSRSPGRRVLDPKDAPVQFRVPPSPACVSG 60  
 QY 61 RAGPHRGNATSFVKQKTIITWMDTKGPTAESSEKNNNTRIDSMSSVQKDNFPHKV 120  
 Db 61 RAGPHRGNATSFVKQKTIITWMDTKGPTAESSEKNNNTRIDSMSSVQKDNFPHKV 120  
 QY 61 RAGPHRGNATSFVKQKTIITWMDTKGPTAESSEKNNNTRIDSMSSVQKDNFPHKV 120  
 Db 61 RAGPHRGNATSFVKQKTIITWMDTKGPTAESSEKNNNTRIDSMSSVQKDNFPHKV 120  
 QY 121 EKLNVPLNLDKSPTEKSSQYLNQQTASVCKWQNEKHAEQLLASEPPAGTLPKQLS 180  
 Db 121 EKLNVPLNLDKSPTEKSSQYLNQQTASVCKWQNEKHAEQLLASEPPAGTLPKQLS 180  
 QY 181 NANIGQSPHTDDHSDTDEEDRDNOQFLTPIKLANTKPTVGGQARSNCCKSGRSQSVK 240  
 Db 181 NANIGQSPHTDDHSDTDEEDRDNOQFLTPIKLANTKPTVGGQARSNCCKSGRSQSVK 240  
 QY 241 CTGCGQEEVDVLPESPLSDVGAEDITGPKNDKLTGQESSLIGDSPPEKESEPEPMDV 300  
 Db 241 CTGCGQEEVDVLPESPLSDVGAEDITGPKNDKLTGQESSLIGDSPPEKESEPEPMDV 300  
 QY 301 DNSRNSQDSSEADETSFVDEQDDRESSQTANKLSSCOAREADGDLRKRYLTGSEVRLH 360  
 Db 301 DNSRNSQDSSEADETSFVDEQDDRESSQTANKLSSCOAREADGDLRKRYLTGSEVRLH 360  
 QY 361 FOEGENAGTDLNAKPSGNSSSLNVECRSSQKHGRDCKITDHPMRISKSEDRKQEC 420  
 Db 361 FOEGENAGTDLNAKPSGNSSSLNVECRSSQKHGRDCKITDHPMRISKSEDRKQEC 420  
 QY 421 EVRHQTERKIPKYPINLPPEKWKLTGPIEMRKMPRCGIHLPISLPSASHTVTVRVLD 480  
 Db 421 EVRHQTERKIPKYPINLPPEKWKLTGPIEMRKMPRCGIHLPISLPSASHTVTVRVLD 480  
 QY 481 LRAGEVPKPPPTHYKDLWKNHKVMPCEQNLNVEDENGERTAGSRWELIQTALLNKFT 540  
 Db 481 LRAGEVPKPPPTHYKDLWKNHKVMPCEQNLNVEDENGERTAGSRWELIQTALLNKFT 540  
 QY 541 RPQNLDAILKYNVAYSKKWDFTALVDFWVKVLEAEQAHLQYSLPDMVKIALCLPNIC 600  
 Db 541 RPQNLDAILKYNVAYSKKWDFTALVDFWVKVLEAEQAHLQYSLPDMVKIALCLPNIC 600  
 QY 601 TOPIPLKQKMHVSVTMSQEQIASLLANAFCTPPRNNAKMSYSSYPDINFNRLFEGR 660  
 Db 601 TOPIPLKQKMHVSVTMSQEQIASLLANAFCTPPRNNAKMSYSSYPDINFNRLFEGR 660  
 QY 661 SSRKPEKLTLCFYPRRVTEKPTGLVTFTRQSLDPEWERCCKPLTRLHVYEGTIEG 720  
 Db 661 SSRKPEKLTLCFYPRRVTEKPTGLVTFTRQSLDPEWERCCKPLTRLHVYEGTIEG 720  
 QY 721 NRGMLQVDFANRFVGGGVTCAGLVQEEIRFLINPELIVSRFLTVELDHNELIITGTEQ 780  
 Db 721 NRGMLQVDFANRFVGGGVTCAGLVQEEIRFLINPELIVSRFLTVELDHNELIITGTEQ 780  
 QY 781 YSEYTGATYRWASHEDGSKDQWQRCTEIVAIIDALHFRYLDQVPEKVRRLNKA 840  
 Db 781 YSEYTGATYRWASHEDGSKDQWQRCTEIVAIIDALHFRYLDQVPEKVRRLNKA 840  
 QY 841 YCGFLRPGVPSNLAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 Db 841 YCGFLRPGVPSNLAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 QY 901 RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECRCNCTPGPDIKLPFIYHVESSEATTD 960  
 Db 901 RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECRCNCTPGPDIKLPFIYHVESSEATTD 960  
 QY 961 MPQQAAGT 968  
 Db 961 MPQQAAGT 968

RESULT 3  
 AAU76013  
 ID AAU76013 standard; protein; 968 AA.  
 XX



Thu May 27 09:56:11 2004

us-09-302-812-6.rag

Db	241	CTCQCEVDVLPESPLSDVGAEDIGTGPKNDKLTQCESSLGDSPPFEKESPEPMDV	300
Qy	301	DNSRNSQDSEADEETSFPVDEQDRSSQTANKLSSCQAREADGLRKRYLTGKSEVRLH	360
Db	301	DNSRNSQDSEADEETSFPVDEQDRSSQTANKLSSCQAREADGLRKRYLTGKSEVRLH	360
Qy	361	FOEGENNAGTSDLNAPSGNSSSLNVECRSSQKHGRKDSKITDHFMRISKSEDRRKEQC	420
Db	361	FOEGENNAGTSDLNAPSGNSSSLNVECRSSQKHGRKDSKITDHFMRISKSEDRRKEQC	420
Qy	421	EVRHQRTERKIPKYPNLPPEKKWLGTPPIEMRMKPRCGIHLPSLPSASHVTYVVDL	480
Db	421	EVRHQRTERKIPKYPNLPPEKKWLGTPPIEMRMKPRCGIHLPSLPSASHVTYVVDL	480
Qy	481	LRAGEVPKPPPTHYKDLWDNKHVMPCEQNLVPEDENGERTAGSRWELIQTALINKFT	540
Db	481	LRAGEVPKPPPTHYKDLWDNKHVMPCEQNLVPEDENGERTAGSRWELIQTALINKFT	540
Qy	541	RPQNLKDALIKYNVAYSKWDFLTALVDFWDXVLEAEAHLYQSILPDMVKIALCLPNIC	600
Db	541	RPQNLKDALIKYNVAYSKWDFLTALVDFWDXVLEAEAHLYQSILPDMVKIALCLPNIC	600
Qy	601	TQPIPLLKQKXNHSVTMSQEQIASLLANAFCTPPRRNAKMSYSSYPDINFNRLPEGR	660
Db	601	TQPIPLLKQKXNHSVTMSQEQIASLLANAFCTPPRRNAKMSYSSYPDINFNRLPEGR	660
Qy	661	SSRPEKULTFCYFRRTYKPKTGLVNTFTQSLDDPPWEKCEKPLRLHVTYEGTTEG	720
Db	661	SSRPEKULTFCYFRRTYKPKTGLVNTFTQSLDDPPWEKCEKPLRLHVTYEGTTEG	720
Qy	721	NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTVLDHNECLITGTEQ	780
Db	721	NGRGLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLFTVLDHNECLITGTEQ	780
Qy	781	YSEYTGATYETYNARSHEDSGKDWQRCTEIVADALHFRYLDQFVPEKVRRELANKA	840
Db	781	YSEYTGATYETYNARSHEDSGKDWQRCTEIVADALHFRYLDQFVPEKVRRELANKA	840
Qy	841	YCGFLRPGVPSNLAVATGNNCGGAFGGDARLKALIQILAAAAAERDVYFTFGDSELM	900
Db	841	YCGFLRPGVPSNLAVATGNNCGGAFGGDARLKALIQILAAAAAERDVYFTFGDSELM	900
Qy	901	RLIYSMTFTLTKLDVGKYLKLLRYNNEECNCSCTPGDKLYPFIYHAVESSAETTD	960
Db	901	RLIYSMTFTLTKLDVGKYLKLLRYNNEECNCSCTPGDKLYPFIYHAVESSAETTD	960
Qy	961	MPGQKAGT 968	
Db	961	MPGQKAGT 968	
RESULT 4			
ID	ABG72280	standard; protein; 968 AA.	
XX	XX		
AC	ABG72280;		
XX	XX		
DT	13-MAR-2003 (first entry)		
DE	Mmurine poly (ADP-ribose) glycohydrolase (PARG) enzyme.		
XX	Murine: poly (ADP-ribose) glycohydrolase; PARG; PARG expression;		
KW	cellular response; DNA damage; neoplastic disorder inducing agent;		
KW	oxidative stress; neoplastic disorder; myocardial infarction;		
KW	vascular stroke; neurodegenerative disorder; Alzheimer's disease;		
KW	Parkinson's disease; Huntington's disease; inborn genetic error;		
KW	reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;		
KW	cystostatic; neuroprotective; nootropic; antiparkinsonian; cardiant;		
KW	vasotrophic; anticonvulsant; cerebroprotective; enzyme.		
XX	XX		
OS	XX	Mus musculus.	
PN	US2002132328-A1.		

XX	PD	19-SEP-2002.	
XX	XX	09-OCT-2001; 2001US-00973451.	
XX	XX	01-MAY-1998; 98US-0083768P.	
XX	XX	30-APR-1999; 99US-00302812.	
XX	XX	(JACO/) JACOBSON M K.	
XX	XX	(JACO/) JACOBSON E L.	
XX	XX	(AMEJ/) AME J.	
XX	XX	(LINW/) LIN W.	
XX	XX	Jacobson MK, Jacobson EL, Ame J, Lin W;	
XX	XX	WPI; 2003-155895/15.	
XX	XX	N-PSDB; ABX14479.	
XX	XX	New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)	
XX	XX	glycohydrolase activity, for preventing, treating, or ameliorating a	
XX	XX	disease condition, e.g. neoplastic disorder, myocardial infarction or	
XX	XX	vascular stroke.	
XX	XX	Claim 28; Fig 16; 86pp; English.	
XX	XX	The present invention relates to the isolation of poly(ADP-ribose)	
XX	XX	glycohydrolase (PARG) from several species, and the polynucleotide	
XX	XX	sequences encoding them. Methods for inhibiting PARG expression or	
XX	XX	overexpressing PARG are also disclosed. PARG is involved in the cellular	
XX	XX	response to DNA damage, and is associated with the body's response to	
XX	XX	neoplastic disorder inducing agents and oxidative stress. The	
XX	XX	polynucleotide sequences encoding PARG and PARG modulators are useful for	
XX	XX	preventing, treating, or ameliorating diseases such as neoplastic	
XX	XX	disorders, myocardial infarction, vascular stroke, neurodegenerative	
XX	XX	diseases (e.g. Alzheimer's disease, Parkinson's disease or Huntington's	
XX	XX	disease), inborn genetic errors, reperfusion following ischaemia, aging,	
XX	XX	and neurotoxicity. The polynucleotide sequences are also useful in gene	
XX	XX	therapy. The methods are useful for identifying an agent that modulates	
XX	XX	PARG activity, identifying a mutant PARG allele in an individual, or	
XX	XX	screening candidate molecules for PARG modulating activity. The present	
XX	XX	sequence represents murine PARG enzyme	
XX	XX	Sequence 968 AA;	
XX	XX	Query Match 100.0%; Score 5164; DB 6; Length 968;	
XX	XX	Best Local Similarity 100.0%; Pred. No. 0;	
XX	XX	Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MSAGPGWEPCTKARWGAAGTSAPTASDSNSPFGRRVLDPKDAPVQFRVPPSSPACVSG	60
Db	1	MSAGPGWEPCTKARWGAAGTSAPTASDSNSPFGRRVLDPKDAPVQFRVPPSSPACVSG	60
Qy	61	RAGPHRGNATSFVKQKTTTWMDTKGPTASESKENNTRIDSMWSSVQKDNFYPHKV	120
Db	61	RAGPHRGNATSFVKQKTTTWMDTKGPTASESKENNTRIDSMWSSVQKDNFYPHKV	120
Qy	121	EKLENVPLNLDKSPTEKSSQYLNQQTATASVCKWQNEKGHAEQALLASEPAGTFLPKQLS	180
Db	121	EKLENVPLNLDKSPTEKSSQYLNQQTATASVCKWQNEKGHAEQALLASEPAGTFLPKQLS	180
Qy	181	NANIGQSPHTDHSDDHEEDRONQOFLPIKLANTKPTVGDGQARSCKSCSRSQSKD	240
Db	181	NANIGQSPHTDHSDDHEEDRONQOFLPIKLANTKPTVGDGQARSCKSCSRSQSKD	240
Qy	241	CTCQCEVDVLPESPLSDVGAEDIGTGPKDNKLTQCESSLGDSPPFEKESPEPMDV	300
Db	241	CTCQCEVDVLPESPLSDVGAEDIGTGPKDNKLTQCESSLGDSPPFEKESPEPMDV	300
Qy	301	DNSRNSQDSEADEETSFPVDEQDRSSQTANKLSSCQAREADGLRKRYLTGSEVRLH	360
Db	301	DNSRNSQDSEADEETSFPVDEQDRSSQTANKLSSCQAREADGLRKRYLTGSEVRLH	360
Qy	361	FOEGENNAGTSDLNAPSGNSSSLNVECRSSQKHGRKDSKITDHFMRISKSEDRRKEQC	420

Thu May 27 09:56:11 2004

Db 361 FQFEGENAGTSDLNAKPSGNSLNVCEKSKQHGKDSKLTDFMISKSDRRKQCC 420  
 Qy 421 EVRHQRTKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSASHTVTVRDL 480  
 Db 421 EVRHQRTKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSASHTVTVRDL 480  
 Qy 481 LRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELIQTALINKFT 540  
 Db 481 LRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELIQTALINKFT 540  
 Qy 541 RPQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVKIALCLPNIC 600  
 Db 541 RPQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVKIALCLPNIC 600  
 Qy 601 TOPIPLKQNMHSVTMSQEQIASLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 Db 601 TOPIPLKQNMHSVTMSQEQIASLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 Qy 661 SSRKPEKLTLCYFRVTEKXPTGLVTFTRQSLDFPWEKCEKPLTLHVTYEGTIG 720  
 Db 661 SSRKPEKLTLCYFRVTEKXPTGLVTFTRQSLDFPWEKCEKPLTLHVTYEGTIG 720  
 Qy 721 NGRGMLQVDFANRFVGGVGTGAGLVOEIRFLINPELIVSRLEFTEVLDHNECLIIITGTEQ 780  
 Db 721 NGRGMLQVDFANRFVGGVGTGAGLVOEIRFLINPELIVSRLEFTEVLDHNECLIIITGTEQ 780  
 Qy 781 YSEYTGVAETRWARSHEDSEKDDQRRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 Db 781 YSEYTGVAETRWARSHEDSEKDDQRRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 Qy 841 YCGFLRPGVPSNLAVATGNWCGAPGPGDARLKALIQILAAAAARDVVYFTFGSELM 900  
 Db 841 YCGFLRPGVPSNLAVATGNWCGAPGPGDARLKALIQILAAAAARDVVYFTFGSELM 900  
 Qy 901 RDIYSMTFTLTKLDVGVKYLRLRYNEECRCNCTPGPDIKLPIYHAYESSAETTD 960  
 Db 901 RDIYSMTFTLTKLDVGVKYLRLRYNEECRCNCTPGPDIKLPIYHAYESSAETTD 960  
 Qy 961 MPQKACT 968  
 Db 961 MPQKACT 968  
 RESULT 5  
 AAE25630  
 ID AAE25630 standard; protein; 976 AA.  
 AC  
 XX  
 XX AAE25630;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Human poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 DE Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytosstatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 KW  
 OS Homo sapiens.  
 XX  
 XX US6395543-B1.  
 FN  
 XX  
 XX 28-MAY-2002.  
 PD  
 XX  
 XX 23-FEB-2000; 2000US-00511507.  
 PF  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 PR  
 XX 30-APR-1999; 99US-00302812.  
 PR  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42082.  
 DR  
 XX  
 PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage, and  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 Claim 3; Col 55-60; 77pp; English.  
 PS The invention relates to an isolated nucleic acid molecule which encodes  
 XX a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (PARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG  
 XX  
 SQ Sequence 976 AA;  
 Query Match 85.1%; Score 4395; DB 5; Length 976;  
 Best Local Similarity 85.7%; Pred. No. 0; Mismatches 10; Gaps 6;  
 Matches 837; Conservative 52;  
 Qy 1 MSAGPCWEPCTKA-RWGAAGTASPTASDSRSPFGRQRRVLDPKDAPVQFVPPSSPACVS 59  
 Db 1 MNAGPCWEPCTKA-RWGAAGTASPTASDSRSPFGRQRRVLDPKDAPVQFVPPSSPACVS 59  
 Qy 60 GRAGPHRGNATSVFVKQKTIITWMDTKGPKTASE---SKENNTRIDSMWSSVQKDNFY 116  
 Db 60 GRAGPHRGNATSVFVKQKTIITWMDTKGPKTASE---SKENNTRIDSMWSSVQKDNFY 116  
 Qy 117 PHKVEKLVNVPQNLNLSKSPTEKSSOYLNQOQTASVCKWQNEGHAEOLLAASEPAGTFLP 176  
 Db 117 PHKVEKLVNVPQNLNLSKSPTEKSSOYLNQOQTASVCKWQNEGHAEOLLAASEPAGTFLP 176  
 Qy 120 QHNVEKLVNYSQSLDKSLTEKSTQYLNOHQTAAMCKWQNEGHAEOLLAASEPAGTFLP 179  
 Db 120 QHNVEKLVNYSQSLDKSLTEKSTQYLNOHQTAAMCKWQNEGHAEOLLAASEPAGTFLP 179  
 Qy 177 KOLSNANIGOSPHTDDHSDTDEEDRDNQOFLPIKLANTKPTVVDGOAR---SNCKCSG 233  
 Db 177 KOLSNANIGOSPHTDDHSDTDEEDRDNQOFLPIKLANTKPTVVDGOAR---SNCKCSG 233  
 Qy 180 EQFSNANIDRSPQNDHSDTDEEDRDNQOFLPIKLANTKPTVVDGOAR---SNCKCSG 239  
 Db 180 EQFSNANIDRSPQNDHSDTDEEDRDNQOFLPIKLANTKPTVVDGOAR---SNCKCSG 239  
 Qy 234 SROSVDKCTCQOEVDVLPESPLSVGAEDIGTPKNDKLTGOESSLGSPPEFEKESE 293  
 Db 234 SROSVDKCTCQOEVDVLPESPLSVGAEDIGTPKNDKLTGOESSLGSPPEFEKESE 293  
 Qy 240 SCHPGEBCACQDEIDVWPSPSLSDVSGEDVGTGSKNDKLTGOESSLGSPPEFEKESE 299  
 Db 240 SCHPGEBCACQDEIDVWPSPSLSDVSGEDVGTGSKNDKLTGOESSLGSPPEFEKESE 299  
 Qy 294 PESMDVDSNRNSCODSEADETSVPFDEODD-RSSQTANKLSSQAREADGLRKYLT 352  
 Db 294 PESMDVDSNRNSCODSEADETSVPFDEODD-RSSQTANKLSSQAREADGLRKYLT 352  
 Qy 300 PESMDVDSNRNSCODSEADETSVPFDEODD-RSSQTANKLSSQAREADGLRKYLT 359  
 Db 300 PESMDVDSNRNSCODSEADETSVPFDEODD-RSSQTANKLSSQAREADGLRKYLT 359  
 Qy 353 KGSVRLHFOPE-GENNAGTSDLNAPSGNSSNLNVECRSKQHGKDSKITDHFMRISK 411  
 Db 353 KGSVRLHFOPE-GENNAGTSDLNAPSGNSSNLNVECRSKQHGKDSKITDHFMRISK 411  
 Qy 360 XGGEVRLHFOPEGGESRTGMNDLNAPSGNSSNLNVECRSKQHGKDSKITDHFMRISK 419  
 Db 360 XGGEVRLHFOPEGGESRTGMNDLNAPSGNSSNLNVECRSKQHGKDSKITDHFMRISK 419  
 Qy 412 SEDRRKQCEVVRHQRTERKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSAS 471  
 Db 412 SEDRRKQCEVVRHQRTERKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSAS 471  
 Qy 420 AEDRRKEQWETKHQRTKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSAN 479  
 Db 420 AEDRRKEQWETKHQRTKIPKYPNNLPPEKMWLGTPIEMRMKMPRCGIIHLPSLRPSAN 479  
 Qy 472 HTVTVRDLIRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELI 531  
 Db 472 HTVTVRDLIRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELI 531  
 Qy 480 HTVTIRVDLIRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELI 539  
 Db 480 HTVTIRVDLIRAGEVPKPPFTHYKDLMDNKHVMKPCSEQNLVPEDENGERTAGSRWELI 539  
 Qy 532 OTALLNKTRFQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVK 591  
 Db 532 OTALLNKTRFQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVK 591  
 Qy 540 QTALLNKTRFQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVK 599  
 Db 540 QTALLNKTRFQNLKDALIKYNVAYSKKWDFTALVDFWKVLEAEAGHLVQSILPDMVK 599

Thu May 27 09:56:11 2004

CC	invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as	
CC	adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma,	
CC	teratocarcinoma, heart attack, stroke, neurodegenerative diseases, neurological disorders including Alzheimer's, Huntington's and Parkinson's diseases,	
CC	and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs.	
CC	Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the human PARG protein of the invention. This protein is one of several PARG proteins	
CC	(AAU76020-AAU76024) of the invention	
XX	Sequence 976 AA;	
XX	Query Match 85.1%; Score 4395; DB 5; Length 976;	
XX	Best Local Similarity 85.7%; Pred. No. 0;	
XX	Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;	
QY	1 MSAGPGWEPCTKA-RWGAAGTSAPTADSRSPFQRRVLDPKDAPVQFRVPPSSPACVS 59	
DB	1 MNAGPGCEPCTKATWGAA-TTSPASDARSFPQRRVLDPKDAHVFQFRVPPSSPACVP 59	
QY	60 GRAGPHRGNAATSFVFKQITITWMDTKGPKTASSE---SKENNTRIDSMSSVQKDNFY 116	
DB	60 GQAGQHGSAATSLVFKQITITWMDTKGPKTASSE---SKENNTRIDSMSSVQKDNFY 119	
QY	117 PKVKEKLENPOLNLDKSPTEKSOYLNOQTASVCKWQNEKGHAEQALLASEPPAGTFLP 176	
DB	120 QNVKELVNVSQLSLDKSLTEKSTQYLNQHTAMCKWQNEKGKTEQLLESEFPQTVTLVP 179	
QY	177 KOLSNANIGOSPHTDDHSDTDHEEDRDNQOPLTPIKLANTKPTVGDQAR---SNCKSG 233	
DB	180 EQFSNANIDRSPQNDHSDTDSENRDNQOPLTIVKLANAKQTTEDEHAREAKHQKSK 239	
QY	234 SRQSVKCTGCOQBEVDVLPESPLSDVGAEDIGTGPKNDNKLTCQESLIGSPDPEKESE 293	
DB	240 SCHGEDCASCQDEIDVWPKSPLSDVGSDEVDGTGSKNDNKLIRQESCLGNSPPPEKESE 299	
QY	294 PESPMVDNNSNCSQDSEADETSPVDEQDD-BSSQATANKLSSCOAREADGDLKRYLT 352	
DB	300 PESPMVDNNSNCSQDSEADETSPVDEQDD-BSSQATANKLSSCOAREADGDLKRYLT 359	
QY	353 KGEVRLHFOPE-GENNAGTSDLNKPSGNSSLNVECRSSQHQKRSKIDTHFMRISK 411	
DB	360 KGEVRLHFOPE-GENNAGTSDLNKPSGNSSLNVECRSSQHQKRSKIDTHFMRISK 419	
QY	412 SEDREKEQCEVRHQRTERKIPKIIPNLPPEKKWLTPIEEMKMPRCGILHPLSPSPAS 471	
DB	420 AEDREKEQWETKHQTERKIPKIIPNLPPEKKWLTPIEEMKMPRCGILHPLSPSAN 479	
QY	472 HTVTVRVLLLRAGEVYKPPPTHYKDLWKNKVKMPCSEONLYPVEDENGERTAGSRWELI 531	
DB	480 HTVTIRVLLLRAGEVYKPPPTHYKDLWKNKVKMPCSEONLYPVEDENGERTAGSRWELI 539	
QY	532 QTALLNKTPQNLDKAILKYNVAYSKKWDFALTVDVFDKVLLEAAQHLQYSLIPDMVK 591	
DB	540 QTALLNKTPQNLDKAILKYNVAYSKKWDFALTVDVFDKVLLEAAQHLQYSLIPDMVK 599	
QY	592 IALCLPNICTQPIPLKQKNHSTVMSQQLASILLANAFCTPPRNNAKMKSEYSSYPDI 651	
DB	600 IALCLPNICTQPIPLKQKNHSTVMSQQLASILLANAFCTPPRNNAKMKSEYSSYPDI 659	
QY	652 NFNLFEGRSSRPEKLTILFCYFRVTEKPKTGLVTFTRQSLDEPPEWERCCKPLTRLH 711	
DB	660 NFNLFEGRSSRPEKLTILFCYFRVTEKPKTGLVTFTRQSLDEPPEWERCCKPLTRLH 719	
QY	712 VTVEGTIEGNRGMQLQVDFANRFVGGVTCAGLVQSEIRFLINPELIVSLFTEVLHDNE 771	
DB	720 VTVEGTIEGNRGMQLQVDFANRFVGGVTCAGLVQSEIRFLINPELIVSLFTEVLHDNE 779	
QY	772 CLITGTGEQSEVTVTAETVWRWASHEDSGEKDQWRCTEIVADALHFRYLDQFVPE 831	
CC		
QY	592 IALCLPNICTQPIPLKQKNHSTVMSQQLASILLANAFCTPPRNNAKMKSEYSSYPDI 651	
DB	600 IALCLPNICTQPIPLKQKNHSTVMSQQLASILLANAFCTPPRNNAKMKSEYSSYPDI 659	
QY	652 NFNLFEGRSSRPEKLTILFCYFRVTEKPKTGLVTFTRQSLDEPPEWERCCKPLTRLH 711	
DB	660 NFNLFEGRSSRPEKLTILFCYFRVTEKPKTGLVTFTRQSLDEPPEWERCCKPLTRLH 719	
QY	712 VTVEGTIEGNRGMQLQVDFANRFVGGVTCAGLVQSEIRFLINPELIVSLFTEVLHDNE 771	
DB	720 VTVEGTIEGNRGMQLQVDFANRFVGGVTCAGLVQSEIRFLINPELIVSLFTEVLHDNE 779	
QY	772 CLITGTGEQSEVTVTAETVWRWASHEDSGEKDQWRCTEIVADALHFRYLDQFVPE 831	
DB	780 CLITGTGEQSEVTVTAETVWRWASHEDSGERDDCERRCTEIVADALHFRYLDQFVPE 839	
QY	832 KVRRELKAYCGFLRPGVSENLSAVATGNWGCAGFGDARLKALIQILAAAAAARDVVY 891	
DB	840 KVRRELKAYCGFLRPGVSENLSAVATGNWGCAGFGDARLKALIQILAAAAAARDVVY 899	
QY	892 FTGDSSELMRDYISMHTFLTERKLDVGKVKYLLRLRYNEECNCGSTPGPDIKLYPFIYHA 951	
DB	900 FTGDSSELMRDYISMHTFLTERKLDVGKVKYLLRLRYNEECNCGSTPGPDIKLYPFIYHA 959	
QY	952 VESSAETDMPGQKAGT 968	
DB	960 VESCAETADHSGQRTGT 976	
DE	RESULT 6	
AAU76021	AAU76021 standard; protein; 976 AA.	
XX	AC AAU76021;	
XX	DT 08-MAY-2002 (first entry)	
XX	Human poly (ADP-ribose) glycohydrolase (PARG) protein sequence.	
XX	Human; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;	
XX	adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;	
XX	leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;	
XX	hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;	
XX	neurodegenerative disease; neurological disorder; Alzheimer's disease;	
XX	Huntington's disease; Parkinson's disease.	
OS	Homo sapiens.	
XX	US6337202-B1.	
XX	08-JAN-2002.	
XX	23-FEB-2000; 2000US-00511477.	
XX	01-MAY-1998; 98US-0083768P.	
XX	30-APR-1999; 99US-00302812.	
XX	(KENT ) UNIV KENTUCKY RES FOUND.	
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;	
XX	WPI; 2002-163240/21.	
XX	N-PSDB; ABK14932.	
XX	Novel isolated and purified poly (ADP-ribose) glycohydrolase protein	
XX	which catalyses release of ADP-ribose from ADP ribose polymer, useful for	
XX	treating neoplastic and neurological disorders, heart attack and stroke.	
XX	Claim 2; Col 55-60; 81pp; English.	
XX	The present invention relates to a new poly (ADP-ribose) glycohydrolase	
XX	(PARG) protein which catalyses release of ADP-ribose from an ADP	
XX	(adenosine diphosphate)-ribose polymer. The PARG molecule of the	



Thu May 27 09:56:11 2004

XX AC ABG72279; DB  
XX XX 13-MAR-2003 (first entry) QY  
XX XX Human poly (ADP-ribose) glycohydrolase (PARG) enzyme. DB  
XX DE  
XX KW Human; poly (ADP-ribose) glycohydrolase; PARG; PARG expression; DB  
KW cellular response; DNA damage; neoplastic disorder inducing agent; DB  
KW oxidative stress; neoplastic disorder; myocardial infarction; DB  
KW vascular stroke; neurodegenerative disorder; Alzheimer's disease; DB  
KW Parkinson's disease; Huntington's disease; inborn genetic error; DB  
KW reperfusion; ischemia; aging; neurotoxicity; PARG activity; cytostatic; DB  
KW neuroprotective; nontropic; antiparkinsonian; cardiac; vasotropic; DB  
KW anticonvulsant; cerebroprotective; enzyme. DB  
XX OS Homo sapiens. DB  
XX XX US2002132328-A1. DB  
XX PN 19-SEP-2002. DB  
XX PD 09-OCT-2001; 2001US-00973451. DB  
XX XX 01-MAY-1998; 98US-0083768P. DB  
XX PR 30-APR-1999; 99US-00302812. DB  
XX XX (JACO/) JACOBSON M K. DB  
XX PA (JACO/) JACOBSON E L. DB  
XX PA (AMEJ/) AME J. DB  
XX PA (LINW/) LIN W. DB  
XX XX Jacobson MK, Jacobson EL, Ame J, Lin W; DB  
XX WPI; 2003-155895/15. DB  
XX DR N-PSDB; ABX14478. DB  
XX XX New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose) DB  
PT glycohydrolase activity, for preventing, treating, or ameliorating a DB  
PT disease condition, e.g. neoplastic disorder, myocardial infarction or DB  
PT vascular stroke. DB  
XX PS Claim 28; Fig 16; 86pp; English. DB  
XX CC The present invention relates to the isolation of poly (ADP-ribose) DB  
CC glycohydrolase (PARG) from several species, and the polynucleotide DB  
CC sequences encoding them. Methods for inhibiting PARG expression or DB  
CC overexpressing PARG are also disclosed. PARG is involved in the cellular DB  
CC response to DNA damage, and is associated with the body's response to DB  
CC neoplastic disorder inducing agents and oxidative stress. The DB  
CC polynucleotide sequences encoding PARG and PARG modulators are useful for DB  
CC preventing, treating, or ameliorating diseases such as neoplastic DB  
CC disorders, myocardial infarction, vascular stroke, neurodegenerative DB  
CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's DB  
CC disease), inborn genetic errors, reperfusion following ischemia, aging, DB  
CC and neurotoxicity. The polynucleotide sequences are also useful in gene DB  
CC therapy. The methods are useful for identifying an agent that modulates DB  
CC PARG activity, identifying a mutant PARG allele in an individual, or DB  
CC screening candidate molecules for PARG modulating activity. The present DB  
XX sequence represents human PARG enzyme DB  
XX SQ Sequence 976 AA; DB

Query Match 85.1%; Score 4395; DB 6; Length 976;  
Best Local Similarity 85.7%; Pred. No. 0;  
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
QY 1 MSAGPGWEPCTKA-RWGAAGTSAFTASDSRFPGRVLDKDAVPQFVBPSSPACVS 59  
DB 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPSRQRRVLDKDAHVQFVBPSSPACVP 59  
QY 60 GRAGPHRGNATSVFKQKTTTMDTKGPKTAESE---SKENNNTRIDSMSSSVQKDNFY 116

60 GQAGQHRGSATSLVFKQKTTTMDTKGPKTAESESLDSKENNNTRIDSMSSSVQKDNFY 119  
117 PHKVEKLENVLPOLNLDKSPTEKSSOYLNQOQASVCKWQNEKGHAEQLLASPPACTPLP 176  
120 QHNVEKLVNVSQSLSDKSLTEKASTQYLNQHQFATAMCKWQNEKGHAEQLLESEPTVTLVP 179  
177 KQLSNANIGCSPHTDHSDTDHEEDRDNQOFTTPIKLANTKTVGDQAR---SNCKCSG 233  
180 EQFSNANIDRSFQNDHSDTDSEENRDNQOFTTVKLANAKQTTDEHAREAKSHQCKSK 239  
234 SRQSVKDCGCOQEVVDVLPESPLSDVGAEDICTGPKNDKLTGQSSSLGDSPPFFKESE 293  
240 SCHPGEDCASCQODEIDVVPKSPSDVGSDEVTGSKNDKLTIRQSSCLGNSPFFKESE 299  
294 PESPMVDNRSNCSQDESEADETSFVDEQDD-RSSQTANKLSSCQAREADGLRKYLT 352  
300 PESPMVDNRSNCSQDESEADETSFVDEQDD-RSSQTANKLSSCQAREADGLRKYLT 359  
353 KGSEVRLHFQFE-GENNAGTSDLNAPSGNSSLNVECSRSSKQHGKDSKITDHPWRISK 411  
360 KGGEVRLHFQFEGGESRTGMNDLNAPLPGNISLNVCSRNSKQHGKDSKITDHLMLPK 419  
412 SEDRRKEQCEVRHQRTERKIPKYPNLPPEKQWLTGPIEEMKMPKRCGHLPSLPSAS 471  
420 AEDRRKEQWETKHQRTERKIPKYPNLPPEKQWLTGPIEEMKMPKRCGHLPSLPSAN 479  
472 HTVTVRVDLLRAGEVPKFPFTHYKDLMDNKHVMPCEQNLYPVEDNGERTAGSRWELI 531  
480 HTVTVRVDLLRAGEVPKFPFTHYKDLMDNKHVMPCEQNLYPVEDNGERTAGSRWELI 539  
532 QTALLNKFTFPQNLKDAILKYNVAYSKKNDFTALVDKQVLEAEAAHLYQSILPDMVK 591  
540 QTALLNKFTFPQNLKDAILKYNVAYSKKNDFTALVDKQVLEAEAAHLYQSILPDMVK 599  
592 IALCLPNICTQPIELLKQKNHVSMTMSQEOIASLLANAFCTFPRRNAMKSEYSSYPDI 651  
600 IALCLPNICTQPIELLKQKNHVSMTMSQEOIASLLANAFCTFPRRNAMKSEYSSYPDI 659  
652 NFNRLFEGRSRKPCKLKTLCFYFRRVTEKKPTGLVTFTRQSLDFPPEWERCEKPLTRH 711  
660 NFNRLFEGRSRKPCKLKTLCFYFRRVTEKKPTGLVTFTRQSLDFPPEWERCEKPLTRH 719  
712 VTYEGTIEGNRGRLQVDFANRFVGGGVTSAGLQVEEIRFLINDELIVSRFTFVLDHNE 771  
720 VTYEGTIEGNRGRLQVDFANRFVGGGVTSAGLQVEEIRFLINDELIVSRFTFVLDHNE 779  
772 CLITGTFOYSEYTCYAEITYWARSHEGSKDQWRCTEIVADALHFRYRLDQFVPE 831  
780 CLITGTFOYSEYTCYAEITYWARSHEGSKDQWRCTEIVADALHFRYRLDQFVPE 839  
832 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVY 891  
840 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVY 899  
892 FTFGSELMDIYSHWTFITERKLDVGVKYLRLRYNEECNCSTPGPDIKLYFIYHA 951  
900 FTFGSELMDIYSHWTFITERKLDVGVKYLRLRYNEECNCSTPGPDIKLYFIYHA 959  
952 VESSAETDMPGQKAGT 968  
960 VESCAETADHSGQRTGT 976

RESULT 9  
AAE25629  
ID AAE25629 standard; protein; 977 AA.  
XX  
AC AAE25629;  
XX AC  
XX DT 04-NOV-2002 (first entry)  
XX DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
XX XX

Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme; ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxicity; cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity; Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac; cycostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy; antisense therapy.

Bos taurus.

US6395543-B1.

28-MAY-2002.

23-FEB-2000; 2000US-00511507.

01-MAY-1998; 98US-0083768P.

30-APR-1999; 99US-00302812.

(KENT ) UNIV KENTUCKY RES FOUND.

Jacobson MK, Jacobson EL, Ame J, Lin W; WPI; 2002-535641/57.

N-PSDB; RAD42081.

New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

Claim 3; Col 47-45; 77pp; English.

The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyzes release of ADP-ribose from an ADP-ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the preparation of antisense therapeutics containing sequences complementary to the mRNA of PARG gene. The antisense therapeutic are useful to treat neoplastic disorders and conditions caused by genotoxic oxidative stress e.g., cardiac disorders, neuronal disorders, reperfusion injury, neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's disease. The invention is useful in gene therapy and antisense therapy.

The present sequence is bPARG

Sequence 977 AA;

Query Match 83.9%; Score 4332.5; DB 5; Length 977;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGPGEPCCK-ARWGAAGTCTASDSSPPGRRVLDPKAPVQFRVPPSSPACVS 59  
DB 1 MSAGPGEPCCKPRDAAATSPASDARSPPGRRVLDSDKAPVQFRVPPSSGCGAL 60

QY 60 GRAGPHRGATSVFVKQITTTWMDTKGPKTAEE---SKENNTRIDSMSSVQKNFY 116  
DB 61 GRAGPHRGATSVFVKQITTTWMDTKGPKTAEE---SKENNTRIDSMSSVQKNFY 120

QY 117 PHVKEKLENNPQNLNDSPPKESQYLNQOQTASVCKWQNEGKHAQQLASEPPAGTPLP 176  
DB 121 QHNEKLENNVSLQGFDSFVEXKTQYLVKHQTAAMCKWQNEGPHSERLLESEPPAVILVP 180

QY 177 KQLSNANIGSPHTDDHSDTHDEEDRNQOFTPIKLANTKPTVGDGQ---ARNCKKSG 233  
DB 181 EQFSNANVDSPPKDDHSDTNSSESDNQOFTLVKLANAKQTMDEQGREARSHQCGK 240

QY 234 SRQSVKCTGCOQEVVDLPESPLSDVGAEDTGTGPKNDKLTGQESSLGSPPPEKESE 293  
DB 241 ACPAEACACQCEETDVSESLSDTGSDEDVGTGLKANRLNRQESSLGSPPPEKESE 300

QY 294 PESPMVDNNSRNSQDSEADETSPVDEQDRESS--QTANKLSSQCAREADGLKRYLT 352

301 PESPMVDNNSRNSQDSEADETSPVDEQDSSAQNTANKPSRFQPREADTELKRS 360  
QY 353 KGEVRLHFQFE-CENWAGTSDLNKAGSGNSLNVECSRSGKHGKDSKITDHFNRISK 411  
DB 361 KGEIRLHFQFEGGESRAGMVDNAGKPGSTSLNVECRNSQHGKDSKITDHFNRISK 420

QY 412 SEDREKQCEVRHORTRKIPKIPPNLPEKQKWLGTPIEEMKMPRCGTHLPSLPAS 471  
DB 421 AEDKXKEQCEMKQRTKIPKIPPNLSPDKKWLGTPIEEMKMPRCGTHLPSLPAS 480

QY 472 HTVTYRVDLIRAGEVPKPFPHYKQNDKHKVMPQSEQNLKYPVEDNGBERTAGSWE 531  
DB 481 HTVTIRVDLLRIGEVKPFPHYKQNDKHKVMPQSEQNLKYPVEDNGBERTAGSWE 540

QY 532 QTALNKFTRPQNLKDALIKYNVAYSKKWDFALVDFWDMKVLBEAQAHLQSYILPMYK 591  
DB 541 QTALNKLTRPQNLKDALIKYNVAYSKKWDFALVDFWDMKVLBEAQAHLQSYILPMYK 600

QY 592 TALCLPNICTQPIELKOKMNSVTMSGOEIASLLANAFCTPPRNAMKSEYSSYDI 651  
DB 601 TALCLPNICTQPIELKOKMNSVTMSGOEIASLLANAFCTPPRNAMKSEYSSYDI 660

QY 652 NFNRLFEGRSRSEKPKLTLCYFRRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTR 711  
DB 661 NFNRLFEGRSRSEKPKLTLCYFRRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTR 720

QY 712 VTVEGTIEGNGRGMLOVDFAFRFVGCGVTGAGLVOBEIRFLINPELIVSRLEPTEVLDHNE 771  
DB 721 VTVEGTIEGNGRGMLOVDFAFRFVGCGVTGAGLVOBEIRFLINPELIVSRLEPTEVLDHNE 780

QY 772 CLITGTQYSEYTGAYETVMAWASHEDSGEKDQWQRCTEIVADILHFRYLDQFVPE 831  
DB 781 CLITGTQYSEYTGAYETVMAWASHEDSGEKDQWQRCTEIVADILHFRYLDQFVPE 840

QY 832 KYRRELKAYCGFLRPGVPSSENLAVATGNWCGAGFGDARLKALQILAAAARDDVY 891  
DB 841 KYRRELKAYCGFLRPGVPSSENLAVATGNWCGAGFGDARLKALQILAAAARDDVY 900

QY 892 FTFGSELMRIYSMHTFLTERKLDVGVKYLKLLRYNEECRNCSTPGDDIKLYPIYHA 951  
DB 901 FTFGSELMRIYSMHTFLTERKLDVGVKYLKLLRYNEECRNCSTPGDDIKLYPIYHA 960

QY 952 VESSAETDMPQKAG 967  
DB 961 VESCTQTINQFGQTG 976

RESULT 10  
AAU76020  
ID AAU76020 standard; protein; 977 AA.  
XX AC AAU76020;  
XX DT 08-MAY-2002 (first entry)  
XX DE Bovine poly (ADP-ribose) glycohydrolase (PARG) protein sequence.  
XX KW Cow; poly (ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose; adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; teratocarcinoma; hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke; neurodegenerative disease; neurological disorder; Alzheimer's disease; Huntington's disease; Parkinson's disease.  
XX OS Bos taurus.  
XX FH Key Location/Qualifiers  
XX FT Region 601..617 /note= "Represents PARG oligopeptide #1"  
XX FT Region 761..770 /note= "Represents PARG oligopeptide #2"  
XX FT Region 771..801 /note= "Represents PARG oligopeptide #2"

Thu May 27 09:56:11 2004

/note= "Represents PARG oligopeptide #3"  
849..880  
/note= "Represents PARG oligopeptide #4"

US6337202-B1.

08-JAN-2002.

23-FEB-2000; 2000US-00511477.

01-MAY-1998; 98US-0083768P.

30-APR-1999; 99US-00302812.

(KENT) UNIV KENTUCKY RES FOUND.

Jacobson MK, Jacobson EL, Ame J, Lin W;

WPI, 2002-163240/21.

N-PsDB; ABL4331.

Novel isolated and purified poly (ADP-ribose) glycohydrolase protein which catalyses release of ADP-ribose from ADP-ribose polymer, useful for treating neoplastic and neurological disorders, heart attack and stroke.

Claim 2; Col 47-52; 81pp; English.

The present invention relates to a new poly(ADP-ribose) glycohydrolase (PARG) protein which catalyses release of ADP-ribose from an ADP (adenosine diphosphate)-ribose polymer. The PARG molecule of the invention is useful for generating antibodies and can be inhibited or activated for diagnosing and treating neoplastic disorders such as adenocarcinoma, leukaemia, lymphoma, melanoma, sarcoma, teratocarcinoma, hyperplasia and hypertrophy, reperfusion following ischaemia, heart attack, stroke, Huntington's and Parkinson's diseases, disorders including Alzheimer's, Huntington's and Parkinson's diseases, and related conditions. PARG levels may be enhanced to suppress DNA repair and increase the cell's susceptibility to chemotherapy drugs. Antagonists of PARG are administered to treat or prevent neoplastic disorders. The present amino acid sequence represents the bovine PARG protein of the invention. This protein is one of several PARG proteins (AAU76020-AAU76024) of the invention

(AAU76020-AAU76024) of the invention

Sequence 977 AA;

Query Match 83.9%; Score 4332.5; DB 5; Length 977;  
Best Local Similarity 84.0%; Pred. No. 0;  
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGFGPEPCTK-ARWGAAGTSAPTASDSRFPGRVLPDPAQVQFRVPPSPACVS 59  
D 1 MSAGFGCEPCTKPRWDAATSPPAASDARSFGRVLPDPAQVQFRVPPSPGCGAL 60  
QY 60 GRAGPHRGNATSFVFKQITTTWMTKPKTAESE---SKENNTRIDSMSSVQKDNFY 116  
D 61 GRAGPHRGNATSLVFKQITTTWMTKPKTAESESKENNTRIDSMSSVQKDNFY 120  
QY 117 PHKVEKLENPQINDKSPKESQVINOQQTASVCKWQNEKGAPQLLASPPAGTLP 176  
D 121 QHNMEKLENVSLQGFDSKSPVEKGTQYLVKQHTAAMCKWQNEGPHSERLLESEPPAVTLVP 180  
QY 177 KOLSNANIGOSPHTDHSDTDHEEDRDNQFLTPKLANTKPTVGGQ---ARSNCKCSG 233  
D 181 EQFSNANVDOSFPKDDHSDTNSSESDNQFLTHVLKLANAKQTMEDQEGREARSHQCKG 240  
QY 234 SROSVDKCTCCQEEVDVLPESPLSDVGAEDIGTGPKNKLGQSSSLGDSPPFKESE 293  
D 241 ACPAEACACQEQEETDWSSEPLSDTGSDEVDGTLGNANLNKRNQSSSLGNSPPFKESE 300  
QY 294 PESPMVDNNSQDSSEADEETSPVDEQDDRSS-QTANKLSSCQAREADGLRKYLT 352  
D 301 PESPMVDNNSQDSSEADEETSPVDEQDDRSSAGTANKPSRFPQREADTELKRESSA 360  
QY 353 KGSEVRLHFQFE-GENNAGTSDLNKAPSGNSSSLNVECRSSKQHGKGRDSKITDHFMRISK 411

Db 361 KGGEIRLHFQFEGRSAGNDVNAKPGSTSLNVECRSSKQHGKGRDSKITDHFMRVPX 420  
QY 412 SEDRRKEQCEVHQRTKIPKYPNLPPEKWKLTGTPIEMERKMPRCGHLPSLPSAS 471  
D 421 AEDRRKEQCEMKTQRTKIPKYPNLPPEKWKLTGTPIEMERKMPRCGHLPSLPSAN 480  
QY 472 HTVTVRVLLRAGEVKKPPFTTHYKDLWDNKHVQWPCSEQNLXPVEDENGERTAGSRWELI 531  
D 481 HTVTVRVLLRAGEVKKPPFTTHYKDLWDNKHVQWPCSEQNLXPVEDENGERTAGSRWELI 540  
QY 532 QTALLNKFTPNLKDAILKYNVAYSKKWDFLTALVDFWDKVLEAEAAQHLYQSILPDMVK 591  
D 541 QTALLNKFTPNLKDAILKYNVAYSKKWDFLTALVDFWDKVLEAEAAQHLYQSILPDMVK 600  
QY 592 IALCLPNICTOPIPLLKQKMNHSVTMSQEQIASILLANAFCTFPRRNKMKSEYSSYPDI 651  
D 601 IALCLPNICTOPIPLLKQKMNHSVTMSQEQIASILLANAFCTFPRRNKMKSEYSSYPDI 660  
QY 652 NFNLFEGRSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTLH 711  
D 661 NFNLFEGRSSRKPEKLTFLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTLH 720  
QY 712 VTYEGTIEGNGMLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLETFEVLHNE 771  
D 721 VTYEGTIEGNGMLQVDFANRFVGGVGTGAGLVQEEIRFLINPELIVSRLETFEVLHNE 780  
QY 772 CLIIITGTEQSYTGYAETIRWARSHEDSKDQRCRTEIVAIDALHFRYLDQFVPE 831  
D 781 CLIIITGTEQSYTGYAETIRWARSHEDSKDQRCRTEIVAIDALHFRYLDQFVPE 840  
QY 832 KVRRELKAYCGFLRPGVSENLSAVATGNCWCGAGFGDARLKALIQILAAAAEDRVVY 891  
D 841 KVRRELKAYCGFLRPGVSENLSAVATGNCWCGAGFGDARLKALIQILAAAAEDRVVY 900  
QY 892 FTFGDSLMRDIIYSMTETETERKLDYGVKVKLLRLYYNEECNRCSTPGPDIKLYPIYHA 951  
D 901 FTFGDSLMRDIIYSMTETETERKLDYGVKVKLLRLYYNEECNRCSTPGPDIKLYPIYHA 960  
QY 952 VESSAETTMPOGQAG 967  
D 961 VESCTQTTPQPGQRTG 976  
RESULT 11  
AAU75799  
ID AAU75799 standard; protein; 977 AA.  
AC AAU75799;  
DT 08-MAY-2002 (first entry)  
XX Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
DE Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
XX adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
KW neurodegeneration; Huntington's disease; Parkinson's disease;  
KW Alzheimer's disease; neurotoxicity.  
XX Bos taurus.  
XX Key Location/Qualifiers  
FH 601..617  
FT Region /note= "Represents PARG oligopeptide #1"  
FT Region /note= "Represents PARG oligopeptide #2"  
FT Region /note= "Represents PARG oligopeptide #3"  
FT Region /note= "Represents PARG oligopeptide #4"  
XX US6333148-B1.



XX	25-DEC-2001.	DB	481	HTVTRVLDLLRIGVFPKPFTHFKDLWNKHVMPCSEQLNLYPVEDENGERRAAGSRWELI	540
XX	30-APR-1999;	QY	532	QTALLNKETRPONLADAILKYNVAYSKKWDFTALVDVFDKWLVEEAQAHLVYOSILPDMVK	591
XX	01-MAY-1998;	DB	541	QTALLNRLTRPONLADAILKYNVAYSKKWDFTALVDVFDKWLVEEAQAHLVYOSILPDMVK	600
XX	(KENT) UNIV KENTUCKY RES FOUND.	QY	592	IACLPNICTQPIPLKQKMHSHVMSQIOIASLLANAFCTFPPRNNAKMKSEYSSYPDI	651
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;	DB	601	IACLPNICTQPIPLKQKMHSHVMSQIOIASLLANAFCTFPPRNNAKMKSEYSSYPDI	660
XX	WPI; 2002-153820/20.	QY	652	NFNRLPEGRSSRKPEKLTLCYFVRVTEKKTGLVTFTRQSLDEDPPEWERCCKPLTRH	711
XX	N-PSDB; ABK14493.	DB	661	NFNRLPEGRSSRKPEKLTLCYFVRVTEKKTGLVTFTRQSLDEDPPEWERCCKPLTRH	720
XX	Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,	QY	712	VTYEGTIEGNRGLQVDFANRFGVGGVTCAGLVQVEIRFLINPELIVSLRFLTEVLDHNE	771
XX	useful potentially for treating diseases associated with DNA damage, e.g.	DB	721	VTYEGTIEGNRGLQVDFANRFGVGGVTCAGLVQVEIRFLINPELIVSLRFLTEVLDHNE	780
XX	cancer.	QY	772	CLIIITGEQYSEYTGAEYRWARSHEDEKDDWRRCTEIVAIIDALHFRYLDQFVPE	831
XX	Claim 3; Col 45-52; 80pp; English.	DB	781	CLIIITGEQYSEYTGAEYRWARSHEDEKDDWRRCTEIVAIIDALHFRYLDQFVPE	840
XX	The present invention relates to a new method for screening compounds for	QY	832	KVRRELKAYCGFLRPGVSENLSAVATGNWCGGAFGCDARLKALIQILAAAEARVYV	891
XX	ability to modulate activity of an enzyme that hydrolyses ADP (adenosine	DB	841	KVRRELKAYCGFLRPGVSENLSAVATGNWCGGAFGCDARLKALIQILAAAEARVYV	900
XX	diphosphate)-ribose from an ADP-ribose polymer. The compounds are	QY	892	FTRGDSLEMDIYSMTFTLTKLVGVNKLRLRYNEECRCNSTPGPDIKLYPFIYHA	951
XX	diphosphate) or activators of PARG (poly(ADP-ribose) glycohydrolase) and	DB	901	FTRGDSLEMDIYSMTFTLTKLVGVNKLRLRYNEECRCNSTPGPDIKLYPFIYHA	960
XX	inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and	QY	952	VESSAETDMPGQKAG 967	
XX	are used to treat or prevent any condition associated with DNA damage,	DB	961	VESSAETDMPGQKAG 976	
XX	e.g. neoplasia, inherited genetic diseases, myocardial infarction,	QY			
XX	e.g. neoplasia, inherited genetic diseases, myocardial infarction,	DB			
XX	vascular stroke, aging and neurodegeneration e.g. Huntington's,	QY			
XX	Parkinson's or Alzheimer's diseases, or neurotoxicity generally.	DB			
XX	Compounds identified by the new method are more effective than known	QY			
XX	inhibitors and have fewer side effects. The present amino acid sequence	DB			
XX	represents the bovine PARG protein of the invention. This protein is one	QY			
XX	of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the	DB			
XX	invention	QY			
XX	Sequence 977 AA;	DB			
XX	Query Match 83.9%; Score 4332.5; DB 5; Length 977;	DB			
XX	Best Local Similarity 84.0%; Pred. No. 0;	DB			
XX	Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;	DB			
QY	1 MSAGGWEPCYK-ARWGAAGTSAPTASDSRSPGQRVLDKADPQVFRVPPSPACVS 59	DB			
DB	1 MSAGGWEPCYK-ARWGAAGTSAPTASDSRSPGQRVLDKADPQVFRVPPSPACVS 60	QY			
QY	60 GRAGPHRGNATSFVFKQKTTTMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116	DB			
DB	61 GRAGPHRGNATSFVFKQKTTTMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 120	QY			
QY	117 PHKVEKLENPQNLNLDKSPTEKSSQVLYNQOQASVCKWQNEGHABQLILASPPAGTPLP 176	DB			
DB	121 QHNMEKLENSQIGFDKSPVEKGTQYLYKHQHTAAMCKWQNEGHABQLILASPPAGTPLP 180	QY			
QY	177 KOLSNANIGOSPHTDHSDTDHEEDRDNOQFLTPKLANTKPTVGGQ---ARSNCKCSG 233	DB			
DB	181 EQFSNANVDQSPKDDHSDTNSSESDNQFLTHVKLANAKQWMEDEQREARSHQKCGK 240	QY			
QY	234 SRSQVDCCTGCOQEEVDVLPESPLSDVGAIDITGPKNDKLTGQESSLGDSPPPEKSE 293	DB			
DB	241 ACPAEACACQEEVDVLPESPLSDVGAIDITGPKNDKLTGQESSLGDSPPPEKSE 300	QY			
QY	294 PESPMVDNRSNCSQDSEADEETSPPVEDEQDRSS-CTANKLSSCOAREADGDLRKYLT 352	DB			
DB	301 PESPMVDNRSNCSQDSEADEETSPPVEDEQDRSS-CTANKLSSCOAREADGDLRKYLT 360	QY			
QY	353 KGSVRLHFOFE-GENNAGTSDLNAPKSGNSSSLNVECRSSKQHGKDSKITDHPMRISK 411	DB			
DB	361 KGSVRLHFOFE-GENNAGTSDLNAPKSGNSSSLNVECRSSKQHGKDSKITDHPMRISK 420	QY			
QY	412 SEDRRKEQCEVRRHQRTERKIPKYPPLNPPKWLGTPIEMRMKMPGCGIHLPSLRPSAS 471	DB			
DB	421 AEDRKEQCEVRRHQRTERKIPKYPPLNPPKWLGTPIEMRMKMPGCGIHLPSLRPSAN 480	QY			
QY	472 HTVTRVLDLLRIGVFPKPFTHFKDLWNKHVMPCSEQLNLYPVEDENGERRAAGSRWELI 531				

RESULT 12	
ABG72278	
ID	ABG72278 standard; protein; 977 AA.
XX	
AC	ABG72278;
DT	13-MAR-2003 (first entry)
XX	
DE	Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.
XX	
KW	Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;
KW	cellular response; DNA damage; neoplastic disorder inducing agent;
KW	oxidative stress; neoplastic disorder; myocardial infarction;
KW	vascular stroke; neurodegenerative disorder; Alzheimer's disease;
KW	Parkinson's disease; Huntington's disease; inborn genetic error;
KW	reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;
KW	neuroprotective; neurotropic; antiparkinsonian; cardiant; vasotropic;
KW	anticonvulsant; cerebroprotective; enzyme.
XX	
OS	Bos taurus.
XX	
PN	US2002132328-A1.
XX	
PD	19-SEP-2002.
XX	
PF	09-OCT-2001; 2001US-00973451.
XX	
PR	01-MAY-1998; 98US-0083768P.
XX	
PR	30-APR-1999; 99US-00302812.
XX	
PA	(JACO/) JACOBSON M K.
XX	
PA	(JACO/) JACOBSON E L.
XX	
PA	(AMEJ/) AME J.
XX	
PI	(LINW/) LIN W.
XX	
PI	Jacobson MK, Jacobson EL, Ame J, Lin W;
XX	



Thu May 27 09:56:11 2004

DR	WPI; 2003-155895/15.	592	IATCLPNTCTQPIPLLKQKNNHVTMSQEQIATSLIANAFCTFFPRNNAKMKSESYSDPI	651
DR	N-PSDB; ABX14477.	601	IATCLPNTCTQPIPLLKQKNNHVTMSQEQIATSLIANAFCTFFPRNNAKMKSESYSDPI	660
XX				
PT	New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)	652	NPNNLFEGRSRKKPEKATLCFYPRRVTEKPKTGLVTTTQSLDFFPEWERCCEKELTRLH	711
PT	glycohydrolase activity, for preventing, treating, or ameliorating a	661	NPNNLFEGRSRKKPEKATLCFYPRRVTEKPKTGLVTTTQSLDFFPEWERCCEKELTRLH	720
PT	disease condition, e.g. neoplastic disorder, myocardial infarction or			
PT	vascular stroke.			
XX				
PS	Claim 28; Fig 16; 86pp; English.	712	VTYEGTIEGNGRGMLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLEFTEVLDHNE	771
XX		721	VTYEGTIEGNGRGMLQVDFANRFVGGVGTAGLVQEEIRFLINPELIVSRLEFTEVLDHNE	780
CC	The present invention relates to the isolation of poly(ADP-ribose)	772	CLIIITGTEQSYTGYAETIRWARSHEDSGEXDKQORCTEIVADALHFRYLDQFVPE	831
CC	glycohydrolase (PARG) from several species, and the polynucleotide			
CC	sequences encoding them. Methods for inhibiting PARG expression or	781	CLIIITGTEQSYTGYAETIRWARSHEDSGEXDKQORCTEIVADALHFRYLDQFVPE	840
CC	overexpressing PARG are also disclosed. PARG is involved in the cellular			
CC	response to DNA damage, and is associated with the body's response to	832	KYRRELKAYCCFLAPGVSENLAVATGNWCGAFGGDARLKALIQILAAAAAERDVVY	891
CC	neoplastic disorder inducing agents and oxidative stress. The			
CC	polynucleotide sequences encoding PARG and PARG modulators are useful for	841	KIRRELKAYCCFLAPGVSENLAVATGNWCGAFGGDARLKALIQILAAAAAERDVVY	900
CC	preventing, treating, or ameliorating diseases such as neoplastic			
CC	disorders, myocardial infarction, vascular stroke, neurodegenerative	892	FTFGSELMDIYSMTFTLTKRLDVKYKGLLLRYNNEECNCSCTPGPDIKLYPIYHA	951
CC	disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's			
CC	disease), inborn genetic errors, reperfusion following ischemia, aging,	901	FTFGSELMDIYSMTFTLTKRLDVKYKGLLLRYNNEECNCSCTPGPDIKLYPIYHA	960
CC	and neurotoxicity. The polynucleotide sequences are also useful in gene			
CC	therapy. The methods are useful for identifying an agent that modulates	952	VSSAETTMPOGKAG 967	
CC	PARG activity, identifying a mutant PARG allele in an individual, or			
CC	screening candidate molecules for PARG modulating activity. The present	961	VESCTQTTPQPGORTG 976	
CC	sequence represents bovine PARG enzyme			
XX				
SQ	Sequence 977 AA;			
Query Match 83.98; Score 4332.5; DB 6; Length 977;				
Best Local Similarity 84.08; Pred. No. 0;				
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;				
QY	1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRSPFQQRRLVDPKADPQVFRVPSSPACVS 59			
DB	1 MSAGPGCEPTKRRWDAAATSPPAASDARSFPQQRRLVDSKADPQVFRVPSSGAL 60			
QY	60 GRAGPHRGNATSFYFKQKTIITWMDTKGPKTAESE---SKENNTRIDSMSSVOKDNFY 116			
DB	61 GRAGQHRGSATSLVFKQKTIITWMDTKGPKTAESESLHSEKNNTRBESNMSSVOKDNFY 120			
QY	117 PHKVEKLENVQLNDKSPTEKSSQYLNQOQTASVCKWQNGKHAQLLASEPPAGTPLP 176			
DB	121 QHNMEKLENSQLGFDKSPVEKGTQYLQHQHTAAWCKWQNGPHSERLLESEPPAVTLVP 180			
QY	177 KOLSNANIGQSPHTDDHSDTDEEDRDNQQLFTPIKLANAKTMEDEQGREARSHQKCGK 233			
DB	181 EQFSNANVDQSSPKDDHSDTNSERSDNOQLFTHVKLANAKTMEDEQGREARSHQKCGK 240			
QY	234 SRQSVKDCGQEQEVVLPSPSLSDVGAEDIGTGPKNDKLTGQESSIGDSPPEKESE 293			
DB	241 ACHPAEACAGCQEQEEDVVSFSLSDTGSDEVTGLKKNANLNROESSLGNSSPPPEKESE 300			
QY	294 PESPMVDNNSRNSCODSEADERTSPVDEQDRSS-QTANKLSSCQAREADGLAKRYLT 352			
DB	301 PESPMVDNNSKNSCQDEADERTSPVDEQDRSSQAQANKPSRFPQPREADTELAKRGS 360			
QY	353 KGEVRLHFQFE-GENNAGTSLNAPKNSGSSLNVECKSSQKHGKRSKTIIDHFWIRSK 411			
DB	361 KGEIRLHFQFEFGESRACGNDVNAKPGSTSLNVECRNSKHGKRSKTIIDHFWIRSK 420			
QY	412 SEDRKEQCEVHRQTERKIPKIPNLPEKKWLTPTIEMRKWPRCGIHLPSLRPAS 471			
DB	421 AEDKRKEQCEMHQTERKIPKIPNLPEKKWLTPTIEMRKWPRCGIHLPSLRPAS 480			
QY	472 HTVTVPDILLRAGEVPPKPTTHYKDLNDKNGVMPCEQNLYPVEDENGERTAGSRWELI 531			
DB	481 HTVTIIVDILLRAGEVPPKPTTHYKDLNDKNGVMPCEQNLYPVEDENGERTAGSRWELI 540			
QY	532 QTALINKFTRPQNLKDALIKYNVAYSKKWDFTALVDFWVKLEBAQHLQYSLIPDMVK 591			
DB	541 QTALLNRLTFPQNLKDALIKYNVAYSKKWDFTALVDFWVKLEBAQHLQYSLIPDMVK 600			
RESULT 13				
ABBS59491				
XX	ID ABBS59491 standard; protein; 768 AA.			
AC	ABBS59491;			
XX	26-MAR-2002 (first entry)			
DT	Drosophila melanogaster polypeptide SEQ ID NO 5265.			
XX	Drosophila; developmental biology; cell signalling; insecticide;			
XX	pharmaceutical.			
XX	Drosophila melanogaster.			
XX	WO200171042-A2.			
XX	27-SEP-2001.			
XX	23-MAR-2001; 2001WO-US009231.			
XX	23-MAR-2000; 2000US-0191637P.			
XX	11-JUL-2000; 2000US-00614150.			
XX	(PEKE ) PE CORP NY.			
XX	Venter JC, Adams M, Li PWD, Myers EW;			
XX	WPI; 2001-656860/75.			
XX	N-PSDB; ABL03594.			
XX	New isolated nucleic acid detection reagent for detecting 1000 or more			
XX	genes from Drosophila and for elucidating cell signalling and cell-cell			
XX	interactions.			
XX	Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.			
XX	The invention relates to an isolated nucleic acid detection reagent			
XX	capable of detecting 1000 or more genes from Drosophila. The invention is			
XX	useful in developmental biology and in elucidating cell signalling and			
XX	cell-cell interactions in higher eukaryotes for the development of			
XX	insecticides, therapeutics and pharmaceutical drugs. The invention			
XX	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA			
XX	sequences (ABL01840-ABL16175) and the encoded proteins (ABBS5773-			

CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 768 AA;

Query Match 20.8%; Score 1074.5; DB 4; Length 768;  
 Best Local Similarity 41.7%; Pred. No. 1.1e-80;  
 Matches 244; Conservative 96; Mismatches 194; Indels 51; Gaps 15;

QY 408 RISSEDRKKEQCEVHQRTERKIPKIPPNLPP--EKKWLGTPIEMKMPRCGHIHLS 465  
 DB 45 RMSKSPDGGISIEETEE-----PENLANSILDSMRGVSMEAIHR-NRQPELEN 93

QY 466 LRP-SASHTVTVRDILLRAGEVKKPPTTHYKOLWKNKHVMPCEQNLVPEVDENGERTA 524  
 DB 94 LPPVTAGNLHRVMYQLPIRETTPRPVYKSPGK--WDSEHVLRLPCAPESKYPRENPDGSTTI 151

QY 525 GSRWELIQTALINKKTRPQNLKDALIKYNVAYSKKWDFTALVDFWKVLEEAQHLQYS 584  
 DB 152 DFRWEMIERALLQPIKTCCELOQAIIISYNTTYRDQWHFALHQLDDELDSETRVFFED 211

QY 585 ILPDMVKIALCLPNICTQPIELLKQKMHNSVTMSQEQIASLLANAFCTFPFRNA-KMKS 643  
 DB 212 LLPRIIRLALRPDLIQSPVPLKHHKNASLSLSQQIISCLLANAFCLTFPRNTLKRS 271

QY 644 EYSSVPDINFNLFEGRSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--LE 695  
 DB 272 EYSTFPDINFNLQSTGPAVLEKLCIMHYFRVVCPTERDASNVPTGVTVFRRSGLPE 331

QY 696 DPEWERCCKPL--TFLHVTYEGTIEGNGRGMLOVDPANRFVGGVYGAGLVQEEIRFLI 753  
 DB 332 HLIDMSQSAAPLDGVDPLHVAEGTIEDEGLGLQVDFAFKYLGGLGVGHGCVQEEIRFVI 391

QY 754 NPGLIYSRLFTVLDHNECLITGTQYSEYTYGAEYRWARSHEDGSKDQWRCTEI 813  
 DB 392 CPPELLVKGKLFTECLRPFEALVMDLRELNKAYIGFVHMVTPPPG-----VATGNWCGGAF 451

QY 814 VAIDALHFRYLDQFVPEKVRRELKAYCGFLR-----PGVSENLASVATGNWCGGAF 867  
 DB 452 VAIDALHFAQSHQYREDLMERELKAYIGFVHMVTPPPG-----VATGNWCGGAF 503

QY 868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHTFLTERKLDVGKVKYLLRY 927  
 DB 28-MAY-2002.

RESULT 14  
 AAE25632  
 ID AAE25632 standard; protein; 768 AA.  
 XX  
 AC AAE25632;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Fruit fly poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 XX  
 KW Poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cyostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisenase therapy; fruit fly.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN US6395543-B1.  
 XX  
 PD 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 PA Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42084.  
 DR  
 DR  
 XX  
 XX  
 PT New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Example 13; Col 71-76; 77pp; English.  
 PS The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutics are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is fruit fly PARG  
 XX  
 SQ Sequence 768 AA;

Query Match 20.8%; Score 1074.5; DB 5; Length 768;  
 Best Local Similarity 41.7%; Pred. No. 1.1e-80;  
 Matches 244; Conservative 96; Mismatches 194; Indels 51; Gaps 15;

QY 408 RISSEDRKKEQCEVHQRTERKIPKIPPNLPP--EKKWLGTPIEMKMPRCGHIHLS 465  
 DB 45 RMSKSPDGGISIEETEE-----PENLANSILDSMRGVSMEAIHR-NRQPELEN 93

QY 466 LRP-SASHTVTVRDILLRAGEVKKPPTTHYKOLWKNKHVMPCEQNLVPEVDENGERTA 524  
 DB 94 LPPVTAGNLHRVMYQLPIRETTPRPVYKSPGK--WDSEHVLRLPCAPESKYPRENPDGSTTI 151

QY 525 GSRWELIQTALINKKTRPQNLKDALIKYNVAYSKKWDFTALVDFWKVLEEAQHLQYS 584  
 DB 152 DFRWEMIERALLQPIKTCCELOQAIIISYNTTYRDQWHFALHQLDDELDSETRVFFED 211

QY 585 ILPDMVKIALCLPNICTQPIELLKQKMHNSVTMSQEQIASLLANAFCTFPFRNA-KMKS 643  
 DB 212 LLPRIIRLALRPDLIQSPVPLKHHKNASLSLSQQIISCLLANAFCLTFPRNTLKRS 271

QY 644 EYSSVPDINFNLFEGRSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--LE 695  
 DB 272 EYSTFPDINFNLQSTGPAVLEKLCIMHYFRVVCPTERDASNVPTGVTVFRRSGLPE 331

QY 696 DPEWERCCKPL--TFLHVTYEGTIEGNGRGMLOVDPANRFVGGVYGAGLVQEEIRFLI 753  
 DB 332 HLIDMSQSAAPLDGVDPLHVAEGTIEDEGLGLQVDFAFKYLGGLGVGHGCVQEEIRFVI 391

QY 754 NPGLIYSRLFTVLDHNECLITGTQYSEYTYGAEYRWARSHEDGSKDQWRCTEI 813  
 DB 392 CPPELLVKGKLFTECLRPFEALVMDLRELNKAYIGFVHMVTPPPG-----VATGNWCGGAF 451

QY 814 VAIDALHFRYLDQFVPEKVRRELKAYCGFLR-----PGVSENLASVATGNWCGGAF 867  
 DB 452 VAIDALHFAQSHQYREDLMERELKAYIGFVHMVTPPPG-----VATGNWCGGAF 503

QY 868 GGDARLKALIQILAAAAAERDVVYFTFGDSELMDIYSMHTFLTERKLDVGKVKYLLRY 927  
 DB 28-MAY-2002.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:38:05 ; Search time 31.0181 Seconds  
(without alignments)  
1611.120 Million cell updates/sec

Title: US-09-302-812-6  
Perfect score: 5164  
Sequence: 1 MSAGPWEPTCKARWGAAGT.....YHVESSAETDMPGQKAGT 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5164	100.0	968	US-09-302-812-6	Sequence 6, Appli
2	5164	100.0	968	US-09-511-477-6	Sequence 6, Appli
3	5164	100.0	968	US-09-511-507-6	Sequence 6, Appli
4	4395	85.1	976	US-09-302-812-4	Sequence 4, Appli
5	4395	85.1	976	US-09-511-477-4	Sequence 4, Appli
6	4395	85.1	976	US-09-511-507-4	Sequence 4, Appli
7	4332.5	83.9	977	US-09-302-812-2	Sequence 2, Appli
8	4332.5	83.9	977	US-09-511-477-2	Sequence 2, Appli
9	4332.5	83.9	977	US-09-511-507-2	Sequence 2, Appli
10	1074.5	20.8	768	US-09-302-812-8	Sequence 8, Appli
11	1074.5	20.8	768	US-09-511-477-8	Sequence 8, Appli
12	1074.5	20.8	768	US-09-511-507-8	Sequence 8, Appli
13	443.5	8.6	726	US-09-302-812-10	Sequence 10, Appli
14	443.5	8.6	726	US-09-511-477-10	Sequence 10, Appli
15	443.5	8.6	726	US-09-511-507-10	Sequence 10, Appli
16	167	3.2	31	US-09-302-812-11	Sequence 11, Appli
17	167	3.2	31	US-09-511-477-11	Sequence 11, Appli
18	167	3.2	31	US-09-511-507-11	Sequence 11, Appli
19	167	3.2	1177	US-09-134-001C-5106	Sequence 5106, Ap
20	163.5	3.2	1187	US-08-320-559-28	Sequence 28, Appli
21	163.5	3.2	1187	US-08-545-860D-28	Sequence 28, Appli
22	163.5	3.2	1187	PCT-US94-04496-28	Sequence 28, Appli
23	163.5	3.2	1210	US-08-320-559-26	Sequence 26, Appli
24	163.5	3.2	1210	US-08-545-860D-26	Sequence 26, Appli
25	163.5	3.2	1210	PCT-US94-04496-26	Sequence 26, Appli
26	153	3.0	1792	US-08-962-284-4	Sequence 4, Appli
27	150	2.9	1115	US-08-568-459A-2	Sequence 2, Appli

28	150	2.9	1115	2	US-08-487-826B-2	Sequence 2, Appli
29	150	2.9	1115	4	US-09-210-288-2	Sequence 2, Appli
30	150	2.9	1115	6	5198347-6	Patent No. 5198347
31	145.5	2.8	559	1	US-08-320-559-31	Sequence 31, Appli
32	145.5	2.8	559	3	US-08-545-860D-31	Sequence 31, Appli
33	145.5	2.8	559	5	PCT-US94-04496-31	Sequence 31, Appli
34	144	2.8	26	4	US-09-302-812-21	Sequence 21, Appli
35	144	2.8	26	4	US-09-511-477-21	Sequence 21, Appli
36	144	2.8	26	4	US-09-511-507-21	Sequence 21, Appli
37	144	2.8	29	4	US-09-302-812-12	Sequence 12, Appli
38	144	2.8	29	4	US-09-511-477-12	Sequence 12, Appli
39	144	2.8	29	4	US-09-511-507-12	Sequence 12, Appli
40	137.5	2.7	913	4	US-08-971-089-4	Sequence 4, Appli
41	137	2.7	1805	1	US-07-853-913-2	Sequence 2, Appli
42	136.5	2.6	1435	2	US-08-568-459A-4	Sequence 4, Appli
43	136.5	2.6	1435	2	US-08-487-826B-4	Sequence 4, Appli
44	136.5	2.6	1435	4	US-09-210-288-4	Sequence 4, Appli
45	136.5	2.6	1848	3	US-08-296-791-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-09-302-812-6  
; Sequence 6, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-302-812-6

Query Match	100.0%;	Score 5164;	DB 4;	Length 968;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	968;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	MSAGPWEPTCKARWGAAGTSA	DSRSGRORRVLDPKADPVOFRVPPSPACVSG	60
Db	1	MSAGPWEPTCKARWGAAGTSA	DSRSGRORRVLDPKADPVOFRVPPSPACVSG	60
QY	61	RAGPHRGNATSFVKOKTIT	TWMDTKGPKTAESKNNNTRIDSMSSVQKONFYPHKV	120
Db	61	RAGPHRGNATSFVKOKTIT	TWMDTKGPKTAESKNNNTRIDSMSSVQKONFYPHKV	120
QY	121	EKLENVQLNLDKSPTEK	SOYLNQOQTASVCKWQNEKGIAEQLLASEPAGTLPKOLS	180
Db	121	EKLENVQLNLDKSPTEK	SOYLNQOQTASVCKWQNEKGIAEQLLASEPAGTLPKOLS	180
QY	181	NANIGQSPHTDDHSDT	THEEDRDNQOFLTPIKLANTKPTVGDGQARNCCKSGRSQSVKD	240
Db	181	NANIGQSPHTDDHSDT	THEEDRDNQOFLTPIKLANTKPTVGDGQARNCCKSGRSQSVKD	240
QY	241	CTGCGQEEVDVLPES	PLSDVGAEDIGTGPKNDNKLTCQESSLGDSPPFEKESEPESEPMDV	300
Db	241	CTGCGQEEVDVLPES	PLSDVGAEDIGTGPKNDNKLTCQESSLGDSPPFEKESEPESEPMDV	300
QY	301	DNSRNSQDSEAD	BETSPVDFEQDDRSQTKANLSSQAREADGLKRYLTGKSEVRLH	360

301 DNRNSCODSEADEETSPVFEQDRSSQATANKLSSCOAREADGLKRYLTKGSEVRLH 360  
 361 FQFEGENNAGTSDLNAPKPSGNSSINVECRSSKHQGRDSDKITHFMRISKSEDRRKEQC 420  
 361 FQFEGENNAGTSDLNAPKPSGNSSINVECRSSKHQGRDSDKITHFMRISKSEDRRKEQC 420  
 421 EVRHQRTERKIPKYPNLPPEKKWLTPTIEMEMKMPRCGTHLPSLPSASHTVTVRVDL 480  
 421 EVRHQRTERKIPKYPNLPPEKKWLTPTIEMEMKMPRCGTHLPSLPSASHTVTVRVDL 480  
 481 LRAGEVPKPPPTHYKDLWDMKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALNKFT 540  
 481 LRAGEVPKPPPTHYKDLWDMKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALNKFT 540  
 541 RPQNLKDALILKYNVAYSKKWDFTALVDFWVKVLEAEAOHLIYQSILPDMVKIALCLPNIC 600  
 541 RPQNLKDALILKYNVAYSKKWDFTALVDFWVKVLEAEAOHLIYQSILPDMVKIALCLPNIC 600  
 601 TQPIPLLLKQNMHNSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 601 TQPIPLLLKQNMHNSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDPPWERCERKPLTRLHVTEGTTIEG 720  
 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDPPWERCERKPLTRLHVTEGTTIEG 720  
 721 NGRGMLQVDFANRFVGGVGTAGLVQVEIRFLINPELIVSRLEFVLDHNECLIIITGTEQ 780  
 721 NGRGMLQVDFANRFVGGVGTAGLVQVEIRFLINPELIVSRLEFVLDHNECLIIITGTEQ 780  
 781 YSEYTGAEITYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 781 YSEYTGAEITYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 841 YCGFLRPGVPSNLSAVATGNWCGAPGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 841 YCGFLRPGVPSNLSAVATGNWCGAPGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECRCNSTPGDIIKLYPFYHAVESSAETTD 960  
 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECRCNSTPGDIIKLYPFYHAVESSAETTD 960  
 961 MPQXKAGT 968  
 961 MPQXKAGT 968

RESULT 2  
 US-09-511-477-6  
 ; Sequence 6, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLTRANSFERASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-511-477-6

Query Match 100.0%; Score 5164; DB 4; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSPFGQRORRVLPDKAPVQFRVPPSPACVSG 60  
 DB 1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSPFGQRORRVLPDKAPVQFRVPPSPACVSG 60  
 QY 61 RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKESKNNNTRIDSMSSVQKONFYPHKV 120  
 DB 61 RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKESKNNNTRIDSMSSVQKONFYPHKV 120  
 QY 121 EKLENVPOINLDSKSPTEKSSQYLNQOQTASVCKWQNEKGKHAQOLLASEPPAGTLPKQLS 180  
 DB 121 EKLENVPOINLDSKSPTEKSSQYLNQOQTASVCKWQNEKGKHAQOLLASEPPAGTLPKQLS 180  
 QY 181 NANIGOSPHTDHSDDTHEDRDNQOFLPTIKLANTKPTVGDGOARSNCKSCSRQSVKD 240  
 DB 181 NANIGOSPHTDHSDDTHEDRDNQOFLPTIKLANTKPTVGDGOARSNCKSCSRQSVKD 240  
 QY 241 CTGCOQEEVDVLPESPLSDVGABDICTGPKNDNKLJTGQESSLGDSPPEFEKSEPEPMOY 300  
 DB 241 CTGCOQEEVDVLPESPLSDVGABDICTGPKNDNKLJTGQESSLGDSPPEFEKSEPEPMOY 300  
 QY 301 DNRNSCODSEADEETSPVFEQDRSSQATANKLSSCOAREADGLKRYLTKGSEVRLH 360  
 DB 301 DNRNSCODSEADEETSPVFEQDRSSQATANKLSSCOAREADGLKRYLTKGSEVRLH 360  
 QY 361 FQFEGENNAGTSDLNAPKPSGNSSINVECRSSKHQGRDSDKITHFMRISKSEDRRKEQC 420  
 DB 361 FQFEGENNAGTSDLNAPKPSGNSSINVECRSSKHQGRDSDKITHFMRISKSEDRRKEQC 420  
 QY 421 EVRHQRTERKIPKYPNLPPEKKWLTPTIEMEMKMPRCGTHLPSLPSASHTVTVRVDL 480  
 DB 421 EVRHQRTERKIPKYPNLPPEKKWLTPTIEMEMKMPRCGTHLPSLPSASHTVTVRVDL 480  
 QY 481 LRAGEVPKPPPTHYKDLWDMKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALNKFT 540  
 DB 481 LRAGEVPKPPPTHYKDLWDMKHKVMPCEQNLYPVEDENGERTAGSRWELIQTALNKFT 540  
 QY 541 RPQNLKDALILKYNVAYSKKWDFTALVDFWVKVLEAEAOHLIYQSILPDMVKIALCLPNIC 600  
 DB 541 RPQNLKDALILKYNVAYSKKWDFTALVDFWVKVLEAEAOHLIYQSILPDMVKIALCLPNIC 600  
 QY 601 TQPIPLLLKQNMHNSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 DB 601 TQPIPLLLKQNMHNSVTMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
 QY 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDPPWERCERKPLTRLHVTEGTTIEG 720  
 DB 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDPPWERCERKPLTRLHVTEGTTIEG 720  
 QY 721 NGRGMLQVDFANRFVGGVGTAGLVQVEIRFLINPELIVSRLEFVLDHNECLIIITGTEQ 780  
 DB 721 NGRGMLQVDFANRFVGGVGTAGLVQVEIRFLINPELIVSRLEFVLDHNECLIIITGTEQ 780  
 QY 781 YSEYTGAEITYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 DB 781 YSEYTGAEITYRWARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 QY 841 YCGFLRPGVPSNLSAVATGNWCGAPGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 DB 841 YCGFLRPGVPSNLSAVATGNWCGAPGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 QY 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECRCNSTPGDIIKLYPFYHAVESSAETTD 960  
 DB 901 RDIYSMHTFLTERKLDVGKVKYKLLRYNNEECRCNSTPGDIIKLYPFYHAVESSAETTD 960  
 QY 961 MPQXKAGT 968  
 DB 961 MPQXKAGT 968

RESULT 3  
 US-09-511-507-6  
 ; Sequence 6, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; FILE REFERENCE: THEREWITH  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; PRIOR FILING DATE: 2000-02-23  
 ; PRIOR FILING DATE: 09/302,812  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-511-507-6  
 Query Match 100.0%; Score 5164; DB 4; Length 968;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 968; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPQRRVLDPKDAPVQFRVPPSPACVSG 60  
 Db 1 MSAGPGWEPCTKARWGAAGTSAPTASDSRSFPQRRVLDPKDAPVQFRVPPSPACVSG 60  
 QY 61 RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKNNTRIDSMSSVQKDNFYPHKV 120  
 Db 61 RAGPHRGNATSFVFKQKTIITWMDTKGPKTAESKNNTRIDSMSSVQKDNFYPHKV 120  
 QY 121 EKLNPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGAEQALLASEPPAGTLPKQLS 180  
 Db 121 EKLNPQLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGAEQALLASEPPAGTLPKQLS 180  
 QY 181 NANIGQSPHTDDSHDTHEDRDNQQLTPIKLANTKPTVGDQARSNCCKSGRSQSVKD 240  
 Db 181 NANIGQSPHTDDSHDTHEDRDNQQLTPIKLANTKPTVGDQARSNCCKSGRSQSVKD 240  
 QY 241 CTGCGQBEVDVLPESPLSDVGAEDIGTGPKNDNKLTCQESSLGDSPPFKESESPMDV 300  
 Db 241 CTGCGQBEVDVLPESPLSDVGAEDIGTGPKNDNKLTCQESSLGDSPPFKESESPMDV 300  
 QY 301 DNGRNSCQDSEADSETSPVDEQDRSSQATANKLSSQAEADGLRKYLTGSEVRLH 360  
 Db 301 DNGRNSCQDSEADSETSPVDEQDRSSQATANKLSSQAEADGLRKYLTGSEVRLH 360  
 QY 361 FQEGENNAGTSDLNAPSGNSSSLNYECSRSSKHGKRDSTKTDHFMRIKSEDRRKEQC 420  
 Db 361 FQEGENNAGTSDLNAPSGNSSSLNYECSRSSKHGKRDSTKTDHFMRIKSEDRRKEQC 420  
 QY 421 EVRHQRTKIPKYPNLPPEKKWLTGPTIEMKMPRCGHIHLSLPSASHTVTVVDL 480  
 Db 421 EVRHQRTKIPKYPNLPPEKKWLTGPTIEMKMPRCGHIHLSLPSASHTVTVVDL 480  
 QY 481 LRAGEVPKPRTHYKDLWDNKHVMPCEQNLYPVEDBNGERTAGSRWELLQIALNKFT 540  
 Db 481 LRAGEVPKPRTHYKDLWDNKHVMPCEQNLYPVEDBNGERTAGSRWELLQIALNKFT 540  
 QY 541 RPQNLKDAILKYNVAYSKKNDFTALVDFWVKVLEAEQAHLQYQILPDMVKIALCLPNIC 600  
 Db 541 RPQNLKDAILKYNVAYSKKNDFTALVDFWVKVLEAEQAHLQYQILPDMVKIALCLPNIC 600  
 QY 601 TQPIPLKQKNHVSVMQEQIASLLANAFCTPFRNAKMSYSSYPDINFNLFEGR 660  
 Db 601 TQPIPLKQKNHVSVMQEQIASLLANAFCTPFRNAKMSYSSYPDINFNLFEGR 660

QY 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDEPWECEKPLTRLHVHYEGTIEG 720  
 Db 661 SSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDEPWECEKPLTRLHVHYEGTIEG 720  
 QY 721 NGRGMLQVDPANRFVGGVGTAGLVQBEIRFLINPELIVSRLEFTEVDHNECLIIITGEQ 780  
 Db 721 NGRGMLQVDPANRFVGGVGTAGLVQBEIRFLINPELIVSRLEFTEVDHNECLIIITGEQ 780  
 QY 781 YSEYTGVAETYWARSHEDSGSKDDWQRCCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 Db 781 YSEYTGVAETYWARSHEDSGSKDDWQRCCTEIVADALHFRYLDQFVPEKVRRELNKA 840  
 QY 841 YCGFLRPGVPSNLSAVATGNCGAGFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 Db 841 YCGFLRPGVPSNLSAVATGNCGAGFGGDARLKALIQILAAAAAERDVVYFTFGDSELM 900  
 QY 901 RDIYSMHTFLTERKLDVGKVKYLLLRYYNEECNCSTPGDPDKLYPIYHAVESSAETTD 960  
 Db 901 RDIYSMHTFLTERKLDVGKVKYLLLRYYNEECNCSTPGDPDKLYPIYHAVESSAETTD 960  
 QY 961 MPOGKAGT 968  
 Db 961 MPOGKAGT 968  
 RESULT 4  
 US-09-302-812-4  
 ; Sequence 4, Application US/09302812B  
 ; Patent No. 6333148  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; FILE REFERENCE: THEREWITH  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 4  
 ; LENGTH: 976  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 US-09-302-812-4  
 Query Match 85.1%; Score 4395; DB 4; Length 976;  
 Best Local Similarity 85.7%; Pred. No. 0;  
 Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
 QY 1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFPQRRVLDPKDAPVQFRVPPSPACVS 59  
 Db 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFPQRRVLDPKDAHVQFRVPPSPACVP 59  
 QY 60 GRAGPHRGNATSFVFKQKTIITWMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116  
 Db 60 GQAGQHRGATSATLVFKQKTIITWMDTKGPKTAESLDSKENNTRIDSMSSVQKDNFY 119  
 QY 117 PHKVEKLENVQLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGAEQALLASEPPAGTLP 176  
 Db 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHTAAMCKWQNEKGKTEQLLESEPTVTLVP 179  
 QY 177 KQLSNANIGQSPHTDDSHDTHEDRDNQQLTPIKLANTKPTVGDQAR---SNCKCSG 233  
 Db 180 EQFSNANIDRSPQNDHSDTDSSENRDQQLTIVKLANAKQTTEDSHAEAKSHQCKSK 239  
 QY 234 SROSVDKCTCQOEEVDVLPESPLSDVGAEDIGTGPKNDNKLTCQESSLGDSPPFKESE 293  
 Db 240 SCHPGEDCASQDDEIDVFPKSPVSDVGSDEVTGSGKNDNKLIRQESCLGNSPFFKESE 299

US-09-511-477-4

QY	294	PESMDVNSRNSCODSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKRYLT	352
Db	300	PESMDVNSKNSCODSEADEETSPVFDEQDDSSQTANKPSRQFQARDADIEFRKRYST	359
QY	353	KGSEVRLHFOPE-GENNAGTSDLNKPSGNSSSLNVECRSKQHGKDSKITDHFMRISK	411
Db	360	KGSEVRLHFOPEGGESRTGMNDLNKLPNGNISLNVECRSKQHGKDSKITDHFMRISK	419
QY	412	SEDRRKEQCEVHQRTERKIPKYPNLPBPKWLGTPIEEMRMPRCGHLHPLSRPAS	471
Db	420	AEDRRKQEWETKHQTERKIPKYPNLPBPKWLGTPIEEMRMPRCGHLHPLSRPAS	479
QY	472	HTVTVRVDLLRAGEVPKFPFTHYKOLMDNKHVKMPCSEONLYPVDENGERTAGSRWELI	531
Db	480	HTVTVRVDLLRAGEVPKFPFTHYKOLMDNKHVKMPCSEONLYPVDENGERTAGSRWELI	539
QY	532	QTALLNKTRPONLKDAILKYNVAYSKKWDFTALVDFWVKLEAEAOHLQSLPDMVK	591
Db	540	QTALLNKTRPONLKDAILKYNVAYSKKWDFTALVDFWVKLEAEAOHLQSLPDMVK	599
QY	592	IALCLPNICTQPIPLLKQKNSHVTMSQEOIASILLANAFCTFPPRNAMKSEYSSYDPI	651
Db	600	IALCLPNICTQPIPLLKQKNSHVTMSQEOIASILLANAFCTFPPRNAMKSEYSSYDPI	659
QY	652	NFNRLFEGRSSRRKPEKLTFCYFRVTEKKTGLVTFTRQSLDFPWERCEKPLTRLH	711
Db	660	NFNRLFEGRSSRRKPEKLTFCYFRVTEKKTGLVTFTRQSLDFPWERCEKPLTRLH	719
QY	712	VTYEGTIEBNGRGMLOVDFANRPFVGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLHDNE	771
Db	720	VTYEGTIEBNGRGMLOVDFANRPFVGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLHDNE	779
QY	772	CLIIITGTEQVSEYTGVAETRYRWARSHEDSGEKDDMORRCTEIVADALHFRYLDQFVPE	831
Db	780	CLIIITGTEQVSEYTGVAETRYRWARSHEDSGEKDDMORRCTEIVADALHFRYLDQFVPE	839
QY	832	KVRRELKAYCGFLRPGVSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVY	891
Db	840	KVRRELKAYCGFLRPGVSENLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVY	899
QY	892	FTFGDSELMRDIYSMTFLTERKLDYGVKVKLLRYNNEECRNCSTPGPDIKLYPIYHA	951
Db	900	FTFGDSELMRDIYSMTFLTERKLDYGVKVKLLRYNNEECRNCSTPGPDIKLYPIYHA	959
QY	952	VESAEETDMPGOKAGT 968	
Db	960	VESCAETADHSGORTGT 976	

RESULT 5  
US-09-511-477-4  
; Sequence 4, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAID 201  
; CURRENT APPLICATION NUMBER: US/09/511,477  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR FILING DATE: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

Query Match	85.1%;	Score 4395;	DB 4;	Length 976;
Best Local Similarity	85.7%;	Pred. No. 0;		
Matches 837; Conservative	52;	Mismatches 78;	Indels 10;	Gaps 6;
QY	1	MSAGPGWEPCTKA-RWGAAGTSAPTASDSRSFGRQRRVLDPKDAPVQFVPPSSPACVS	59	
Db	1	MNAGPGCEPCTKATRWGAA-TTSPAASDARSPPSRVLPDPDAHVQFVPPSSPACVP	59	
QY	60	GRAGPHRGNATSFVKOKLTITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY	116	
Db	60	GQAGQHRGATSILVFQKLTITWMDTKGPKTAESSELSKNNNTRIEWMSSVQKDNFY	119	
QY	117	PHKVEKLENVQNLNLDKSFTEKSSOVINQQQTASVCKWQNEKGHAEQLLASPPAGTPLP	176	
Db	120	QHNVEKLVNVQSLSLDKSLTEKSTQYLNQHQTAAMCKWQNEKGHTEQLLESEPPQVTLVP	179	
QY	177	KQLSNANIGOSPHDDHSDTDHEEDRDNOOFLTPIKLANTKTVGDQAR---SNCKCSG	233	
Db	180	EQFSNANIDRSPQNDHSDTDSENRDNQOFLTPIVKLANAKQTTDEHAREAKSHQCKSK	239	
QY	234	SROSVDKTCQOQEEVDVLPESPLSDVAGARDIGTPNDNKLTCQESSLGDSPPPEKESE	293	
Db	240	SCHFGEDCASQOQDEIDVVPKPSPLSDVAGSEDDVGTGKNDNKLIRQESCLGNSPPPEKESE	299	
QY	294	PESPMVDVNSRNSCODSEADEETSPVFDEQDD-RSSQTANKLSSCQAREADGDLRKLYLT	352	
Db	300	PESPMVDVNSKNSCODSEADEETSPGDFQEDGSSSOTANKPSRFQARDADIEFRKRYST	359	
QY	353	KGSEVRLHFOPE-GENNAGTSDLNKPSGNSSSLNVECRSSKHGKDSKITDHFMRISK	411	
Db	360	KGSEVRLHFOPEGESRTGMNDLNKLPNGNISLNVECRNSKHGKDSKITDHLMLRFX	419	
QY	412	SEDRRKEQCEVHQRTERKIPKYPNLPBPKWLGTPIEEMRMPRCGHLPLSRPAS	471	
Db	420	AEDRRKQEWETKHQTERKIPKYPVPHLSPPDKWLGTPIEEMRMPRCGIRLPLLRFSAN	479	
QY	472	HTVTVRVDLLRAGEVPKFPFTHYKDLMDNKHVKMPCSEONLYPVEDENGERTAGSRWELI	531	
Db	480	HTVTVRVDLLRAGEVPKFPFTHYKDLMDNKHVKMPCSEONLYPVEDENGERTAGSRWELI	539	
QY	532	QTALLNKTRPONLKDAILKYNVAYSKKWDF TALVDFWDKVLAEAAQHLQSLPDMVK	591	
Db	540	QTALLNKTRPONLKDAILKYNVAYSKKWDF TALIDFWDKVLAEAAQHLQSLPDMVK	599	
QY	592	IALCLPNICTQPIPLLKQKNSHVTMSQEOIASILANAFCTFPPRNAKMKSEYSSYDPI	651	
Db	600	IALCLPNICTQPIPLLKQKNSHVTMSQEOIASILANAFCTFPPRNAKMKSEYSSYDPI	659	
QY	652	NFNRLFEGSRSRKPEKLTFLCYFRVTEKKTGLVTFTRQSLDFPWERCEKPLTRLH	711	
Db	660	NFNRLFEGSRSRKPEKLTFLCYFRVTEKKTGLVTFTRQSLDFPWERCEKPLTRLH	719	
QY	712	VTYEGTIEBNGRGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLHDNE	771	
Db	720	VTYEGTIEBNGRGMLOVDFANRFVGGVTSAGLVQEEIRFLINPELIVSRLEFTEVLHDNE	779	
QY	772	CLIIITGTEQSEYTTGVAETRYRWARSHEDSGEKDDMORRCTEIVADALHFRYLDQFVPE	831	
Db	780	CLIIITGTEQSEYTTGVAETRYRWARSHEDSGERDDCERRCTEIVADALHFRYLDQFVPE	839	
QY	832	KVRRELKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVY	891	
Db	840	KVRRELKAYCGFLRPGVPSNLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVY	899	
QY	892	FTFGDSELMRDIYSMTFLTERKLDYGVKVKLLFLRYNNEECRNCSTPGPDIKLYPIYHA	951	
Db	900	FTFGDSELMRDIYSMTFLTERKLDYGVKVKLLFLRYNNEECRNCSTPGPDIKLYPIYHA	959	
QY	952	VESAEETDMPGQAGT 968		
Db	960	VESCAETADHSGORTGT 976		

```

RESULT 6
US-09-511-507-4
; Sequence 4, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-507-4

Query Match      85.1%; Score 4395; DB 4; Length 976;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MSAGPGWEPCTKA-RWGAAGTSAPTASDSRFPQRRVLDPKDAPVQFRVPPSPACVS 59
DB 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSFQRRVLDPKDAPVQFRVPPSPACVP 59
QY 60 GRAGPHRGNATSFVKQKTIITWMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116
DB 60 GOAGQHRGSATSLVFKQKTIITWMDTKGPKTAESLDSKNNNTRIESMMSSVQKDNFY 119
QY 117 PHKVEKLENVQMLNDSPTKSSQYLNQOQTASVCKWQNEGHAKOLLASEPPAGTFLP 176
DB 120 QHNVEKLVNVSQSLDSKLTSEKSYQYLNQHTAAMCKWQNEGHAKOLLASEPPAGTFLP 179
QY 177 KQLSNANIGOSPHTDSDTHDEDRDNOQFLTIPIKLANTKPTVGDQAR---SNCKCSG 233
DB 180 EQFSNANIDRSPQNDHSDTSEENRDNQFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239
QY 234 SROSVDKCTCQBEVDVLPESPLSDVGAEDIGTPKNDNKLTCQESLSDGSPPEKSE 293
DB 240 SCHPGEDCASQODEIDVFPKPLSDVGSDEVDGTSKNDNKLIRQESCLGNSPPPEKSE 299
QY 294 PESPMVDNRSNCQDSEADEETSPVFDEQDD-RSSOTANKLASCOAREADGLRKYLT 352
DB 300 PESPMVDNRSNCQDSEADEETSPVFDEQDDSSQTANKPRFQARDADIEFRKRYST 359
QY 353 KGSEVRLHFOFE-GENNAGTSDLNAPSGNSSSLINVECRSSKHQKSDSKITDHFMRISK 411
DB 360 KGCEVRLHFOFEGESRGTGMDLNALPGLNISLNVCRNSKHQKSDSKITDHLMLPK 419
QY 412 SEDRREKQCEVRRQRTKIPKYPNLPPEKKWLGTPIEEMRWPCRGHILPSLPSAS 471
DB 420 AEDRRKEQWETKQRTKIPKYPVPHLSPDKKWLGTPIEEMRWPCRGHILPSLPSAN 479
QY 472 HTVTVRVLLRAGEVPKPFPTHYKQLDWNGHVMPQCEQNLYPVEDENGERTAGSWELI 531
DB 480 HTVTIRVLLRAGEVPKPFPTHYKQLDWNGHVMPQCEQNLYPVEDENGERTAGSWELI 539
QY 532 QTALLNKFTRPQNLKAILKYNVAYSKKWDFDALVDFWPKVLEAEAAQHLQIILPDMVK 591
DB 540 QTALLNKFTRPQNLKAILKYNVAYSKKWDFDALVDFWPKVLEAEAAQHLQIILPDMVK 599
QY 592 IALCLPNICTOPTLLKQKNHVSITMSQEOIASLANAFAFFCTFFRRNAKMKSEYSYPDI 651

```

```

DB 600 IALCLPNICTOPTLLKQKNHVSITMSQEOIASLANAFAFFCTFFRRNAKMKSEYSYPDI 659
QY 652 NFNLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKPLTRLH 711
DB 660 NFNLFEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDFPPEWERCEKPLTRLH 719
QY 712 VTYEGTLEGNGRMLOVDFAFRFVGGVGTGAGLVQBEIRFLINDELIVSRILFTEVLDFNE 771
DB 720 VTYEGTLEENGQMLQVDFAFRFVGGVGTGAGLVQBEIRFLINDELIVSRILFTEVLDFNE 779
QY 772 CLIIITGEQXSEYTGAEFTYRWARSHEDSGSKDQWRCTEIVDAIDALHFRYRLDQFVPE 831
DB 780 CLIIITGEQXSEYTGAEFTYRWARSHEDSGSKDQWRCTEIVDAIDALHFRYRLDQFVPE 839
QY 832 KYREELNKAFCFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVY 891
DB 840 KORRELNKAFCFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVY 899
QY 892 FTFGDSSELMRDYISMWHTFLTERKLDVGKVKLLARYNEECRCNSTGCPDILKLYPFIYHA 951
DB 900 FTFGDSSELMRDYISMWHTFLTERKLDVGKVKLLARYNEECRCNSTGCPDILKLYPFIYHA 959
QY 952 VESSAETTDMPGQKAGT 968
DB 960 VESCAETADHSGQRTGT 976

RESULT 7
US-09-302-812-2
; Sequence 2, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-302-812-2

Query Match      83.9%; Score 4332.5; DB 4; Length 977;
Best Local Similarity 84.0%; Pred. No. 0;
Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;

QY 1 MSAGPGWEPCTK-ARWGAAGTSAPTASDSRFPQRRVLDPKDAPVQFRVPPSPACVS 59
DB 1 MSAGPGCEPCTKPRWDAAATSPPAASDARSFQRRVLDSDAPVQFRVPPSPSGCAL 60
QY 60 GRAGPHRGNATSFVKQKTIITWMDTKGPKTAESE---SKENNTRIDSMSSVQKDNFY 116
DB 61 GRAGQHRGSATSLVFKQKTIITWMDTKGPKTAESLHSESLHSENNNTRIESMMSSVQKDNFY 120
QY 117 PHKVEKLENVQMLNDSPTKSSQYLNQOQTASVCKWQNEGHAKOLLASEPPAGTFLP 176
DB 121 QHNVEKLVNVSQSLGDKSPVEKGTQYLNQHTAAMCKWQNEGHAKOLLASEPPAGTFLP 180
QY 177 KQLSNANIGOSPHTDSDTHDEDRDNOQFLTIPIKLANTKPTVGDQAR---ARSNCKCSG 233
DB 181 EQFSNANVDQSSPKDHDSDTSEENRDNQFLTHVKLANAKQTMEDQEGREARSHQKCGK 240
QY 234 SRQSVKDCCTGCQOEEVDVLPESPLSDVGAEDIGTPKNDNKLTCQESSLSDGSPPEKSE 293

```







QY 644 EYSSYPDINFNRLFEGRSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--LE 695  
 Db 272 EYSTFPDINFNRLYQSTGPAVLEKLCIMHVFRRVCPTERDASNVPTGVTVFRRSGLPE 331  
 QY 696 DPEWERCEKPL--TRLHVTYEGTIEGNGRGLQVDFANRFVGGVGTGAGLVQOEIRFLI 753  
 Db 332 HLIDWSQSAAPLGDVPLHVAEGTIEDEGIGLQVDFANKYLGGLGVCVQOEIRFVI 391  
 QY 754 NPELIVSRFLTEVLHDNECLITITGEQYSEYTGATYRWARSHEDGSEKQDWQRCTEI 813  
 Db 392 CPELLVGLKFTCLRPFEALVMLGAERYSNYTGAGSPWNGFNEDSTPRDSSGRQRTAI 451  
 QY 814 VAIDALHFRYLDQVPEKVRRELKAYCGFLR-----PGVPSNLAVATGNWGCAGF 867  
 Db 452 VAIDALHFAQSHHQYREDLMERELKAYIGFVHMVTPPG-----VATGNWGCAGF 503  
 QY 868 GGDARLKALIQILAAAAERDVVYTFGDSLMRDYISMHTFLTERKLDVGKVKYLLRY 927  
 Db 504 GDSYKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMMLFRNDGTTVQQLWS-ILRS 562  
 QY 928 YNEECRNCSTPGP-----DIKLYPFIYHAVESSAETTDMPGQA 966  
 Db 563 YSRLIKEKSKPRENKASKKKLYDFI---KEELKKVRDVPGEA 604

RESULT 11  
 US-09-511-477-8  
 ; Sequence 8, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; PRIOR FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-511-477-8

Query Match 20.8%; Score 1074.5; DB 4; Length 768;  
 Best Local Similarity 41.7%; Pred. No. 6.9e-96;  
 Matches 244; Conservative 96; Mismatches 194; Indels 51; Gaps 15;  
 QY 408 RISKSEDRRKEQCEVRHQTERKIPKYIPNLP--EKKWLGTPIEEMKMPRCGIHLPS 465  
 Db 45 RMSKSPDGGISEITEE-----PENLANSLDDSWRGVSMEALHR-NRQPFLEN 93  
 QY 466 LRP-SASHTVTVRVLLRAGEVPKPFTHYKDLWNKHVKMPCSEQNLYPYVEDENGERTA 524  
 Db 94 LPPVTAGNLHRVMYQLPIRETTPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGTTI 151  
 QY 525 GSRWELIQTALLNKFTRPONLKDAILKNVAYSKKWDFTALVDFWVKVLEBAEAQHYQS 584  
 Db 152 DFRWEMIERALLQPIKTCBELQAAIISYNTYRDQWHFRALHQLLDEELDESETRVFFED 211  
 QY 585 ILPDWVKIALCLPNICTQPIPLKQKMHVSMTSQEOIASLLANAFCTFFPRNA-KMKS 643  
 Db 212 LLPRIIRLALRPDLIQSPVPLKHKHKNASLSQQOISCLLANAFLCTFFPRNTLKRS 271  
 QY 644 EYSSYPDINFNRLFEGRSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--LE 695  
 Db 272 EYSTFPDINFNRLYQSTGPAVLEKLCIMHVFRRVCPTERDASNVPTGVTVFRRSGLPE 331  
 QY 696 DPEWERCEKPL--TRLHVTYEGTIEGNGRGLQVDFANRFVGGVGTGAGLVQOEIRFLI 753

Db 272 EYSTFPDINFNRLYQSTGPAVLEKLCIMHVFRRVCPTERDASNVPTGVTVFRRSGLPE 331  
 QY 696 DPEWERCEKPL--TRLHVTYEGTIEGNGRGLQVDFANRFVGGVGTGAGLVQOEIRFLI 753  
 Db 332 HLIDWSQSAAPLGDVPLHVAEGTIEDEGIGLQVDFANKYLGGLGVCVQOEIRFVI 391  
 QY 754 NPELIVSRFLTEVLHDNECLITITGEQYSEYTGATYRWARSHEDGSEKQDWQRCTEI 813  
 Db 392 CPELLVGLKFTCLRPFEALVMLGAERYSNYTGAGSPWNGFNEDSTPRDSSGRQRTAI 451  
 QY 814 VAIDALHFRYLDQVPEKVRRELKAYCGFLR-----PGVPSNLAVATGNWGCAGF 867  
 Db 452 VAIDALHFAQSHHQYREDLMERELKAYIGFVHMVTPPG-----VATGNWGCAGF 503  
 QY 868 GGDARLKALIQILAAAAERDVVYTFGDSLMRDYISMHTFLTERKLDVGKVKYLLRY 927  
 Db 504 GDSYKALLQLMVCAQLGRPLAYYTFGNVEFRDDFHEMMLFRNDGTTVQQLWS-ILRS 562  
 QY 928 YNEECRNCSTPGP-----DIKLYPFIYHAVESSAETTDMPGQA 966  
 Db 563 YSRLIKEKSKPRENKASKKKLYDFI---KEELKKVRDVPGEA 604

RESULT 12  
 US-09-511-507-8  
 ; Sequence 8, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EP  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR FILING DATE: 09/302,812  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-511-507-8

Query Match 20.8%; Score 1074.5; DB 4; Length 768;  
 Best Local Similarity 41.7%; Pred. No. 6.9e-96;  
 Matches 244; Conservative 96; Mismatches 194; Indels 51; Gaps 15;  
 QY 408 RISKSEDRRKEQCEVRHQTERKIPKYIPNLP--EKKWLGTPIEEMKMPRCGIHLPS 465  
 Db 45 RMSKSPDGGISEITEE-----PENLANSLDDSWRGVSMEALHR-NRQPFLEN 93  
 QY 466 LRP-SASHTVTVRVLLRAGEVPKPFTHYKDLWNKHVKMPCSEQNLYPYVEDENGERTA 524  
 Db 94 LPPVTAGNLHRVMYQLPIRETTPRPYKSPGK--WDSEHVRLPCAPESKYPRENPDGTTI 151  
 QY 525 GSRWELIQTALLNKFTRPONLKDAILKNVAYSKKWDFTALVDFWVKVLEBAEAQHYQS 584  
 Db 152 DFRWEMIERALLQPIKTCBELQAAIISYNTYRDQWHFRALHQLLDEELDESETRVFFED 211  
 QY 585 ILPDWVKIALCLPNICTQPIPLKQKMHVSMTSQEOIASLLANAFCTFFPRNA-KMKS 643  
 Db 212 LLPRIIRLALRPDLIQSPVPLKHKHKNASLSQQOISCLLANAFLCTFFPRNTLKRS 271  
 QY 644 EYSSYPDINFNRLFEGRSRKPEKLTLCYFRRV--TEKK-----PTGLVTFTRQS--LE 695  
 Db 272 EYSTFPDINFNRLYQSTGPAVLEKLCIMHVFRRVCPTERDASNVPTGVTVFRRSGLPE 331  
 QY 696 DPEWERCEKPL--TRLHVTYEGTIEGNGRGLQVDFANRFVGGVGTGAGLVQOEIRFLI 753

Db 332 HLIDMSQSAAPLGDVPLHDAEGTIEGIGLLQVDFANKYLGGLVGHGCVQBEIRFVI 391  
 QY 754 NPELIVSRLEFVDHNECLITTEQYSEYTGVAETRNARSHEDSGKDDMORRCTEI 813  
 Db 392 CPELLVGLFTECLRPPEALVMLGAERYSNYTGAGSFENSGNEDSTPRDSSGRRQTAI 451  
 QY 814 VALDALHFRVLDQVPEKVRRELKAYCGFLR-----PCVPSENLSAVATGNWGCAP 867  
 Db 452 VALDALHFAQHGYRDLMERELKAYIGFVHMVMTTPPG-----VATGNWGCAP 503  
 QY 868 GGDARLALQILAAAAAERDVVYFTFGDSFLMRDIYSMTFLTERKLDVGKVKVLLRY 927  
 Db 504 GGDYALKALQLMWCAQLGRPLAYYTGNGVEFRDDFHEMILLFENDGTTVQQLWS-ILRS 562  
 QY 928 YNEECNCSPTGP-----DIKLYPFIYHVESAEATDMPGOKA 966  
 Db 563 YSLRIKESKEPRENKASKKLYDFI---KEELKKVRDVEGGA 604

RESULT 13  
 US-09-302-812-10  
 ; Sequence 10, Application US/09302812B  
 ; Patent No. 6333148  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 10  
 ; LENGTH: 726  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE:  
 US-09-302-812-10

Query Match 8.6%; Score 443.5; DB 4; Length 726;  
 Best Local Similarity 21.9%; Pred. No. 5.2e-34;  
 Matches 184; Conservative 124; Mismatches 278; Indels 255; Gaps 32;

QY 199 EEDRDNQOFLPIKANTKPTVGDGQARSCKSGRSQVKTCTGCGQBEVDVLPESPLS 258  
 Db 14 QDEKDYEDIV-GVGFAHQVPTM---KRRKLTGHGNTTESKEDPE--EPKSRDVFVSSQS 67  
 QY 259 DVGAEDIGTGPK-----NDNKLGTQESSLGD---SPPFKEK 292  
 Db 68 DESQEDSAENPEIAKEVSENCENTLTAKISNIESLDNVTERSEHTLDNHKSTPEMEDV 127  
 QY 293 EPESPMDV-----DNRNSCODSEADETSVPF-DEQD----- 324  
 Db 128 NKNKSNIDVAINSEDEDELVEENKEMRDEGQVQQLSQDLFADQDQLIEYPIGIMKDTTQ 187  
 QY 325 ----DRSSQTANKLSQCARADGLRKRKRYLTGSEVRLHPOFEGBNAGTSDLNKPSG 380  
 Db 188 LQITDSEVTAQKMEIETEDS-----TFVGEDSKATKVTTSSSS 230  
 QY 381 NSSSLNVECRSKQHGKR--DSKITDHFMRISKSE-----DRRKEQCEVRHQ 426  
 Db 231 FLSTVST-CEAPAKGRARYQKELEKHVIAFTEGNLTLPDLNKKVDPDRNYRYCTI---- 285  
 QY 427 TERKTPKYPNLPPEKKWLGTFIEEMKMPRCGHLPSRFSASHTVTVRVLLRAGEV 486  
 Db 286 -----PNFPASQ---GKLRDNRYPGKIVL----- 307

QY 487 PKPPTTHYKOLDNKNKVMPCSEONLYPVEDENGER-----TAG-- 525  
 Db 308 ----PORWREF-----DSRRRDSYFYFKRKLDGLYLCYKTKTYGF 344  
 QY 526 -----SRVELIQTALLNKFTRPQNLKDALIKXNVA-----YSKKWDFTFALVDFWVKVL 573  
 Db 345 MFVGLLHNMWEP-----DPDITYKLPALEMYKEMSELVGREEVLEKFA 388  
 QY 574 EEAQAQHYOSILPDMVKIALCLPNICTQPIPLLKQMMHNSVTMSQEQIASILLANAFCT 633  
 Db 389 RVARIAKTAEDILPE--RIYRLVGDV-----ESATLSHKQCAALVARMFE-- 431  
 QY 634 FPRNAKMGSEYSYDPDINFNRLPEGRSSRPKPKLTKLFCYFRFVTEKKPTGLVTF--TR 691  
 Db 432 ----ARPDSPFS-----FCRILSSDKSICVEKLFELTYFDKMSMDPPDGAVSFLTK 480  
 QY 692 QSLERF-PEW-ERCEKPLTELHVTYGTIRGNRGMLOVDNFANRFVGGVGTGAGLVQEEI 749  
 Db 481 MDKDTFNEEKWKDKLRSLEPEVEFDEMLIEDTAL-CTQVDFAHEHLGGVNLHSGSVQEEI 539  
 QY 750 RFLINPELIVSRLEFTEVLHNECLITITGTQYSEYTGVAETRNARSHEDSGSE-K 803  
 Db 540 RFLMCPMMVMVGMMLCEKMKQLEAISIVGAVVFSYTYGHTLAKWAELOPNHRSQNTNEER 599  
 QY 804 DMQRRCRTEIVAIDALHFR-----YLDQVPKVRRELKAYCGFLRPGVPSENLSAVA 858  
 Db 600 DRFGLRVRTIAIDALFKGSKLDCQTEQLNKANIIREMKKASIGFMSQGPFTNI-PIV 658  
 QY 859 TGNWGCAGGADARLALQILAAAAAERDVVYFTFGDSFLMRDIYSMTFLTERKLDVG 918  
 Db 659 TGNWGCAGFNGDKPLAFILQVIAAGVADRPLHPCSGPEPELAACKKIIERMKQKDTVLG 718  
 QY 919 K 919  
 Db 719 K 719

RESULT 14  
 US-09-511-477-10  
 ; Sequence 10, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 10  
 ; LENGTH: 726  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 ; FEATURE:  
 US-09-511-477-10

Query Match 8.6%; Score 443.5; DB 4; Length 726;  
 Best Local Similarity 21.9%; Pred. No. 5.2e-34;  
 Matches 184; Conservative 124; Mismatches 278; Indels 255; Gaps 32;

QY 199 EEDRDNQOFLPIKANTKPTVGDGQARSCKSGRSQVKTCTGCGQBEVDVLPESPLS 258  
 Db 14 QDEKDYEDIV-GVGFAHQVPTM---KRRKLTGHGNTTESKEDPE--EPKSRDVFVSSQS 67  
 QY 259 DVGAEDIGTGPK-----NDNKLGTQESSLGD---SPPFKEK 292  
 Db 68 DESQEDSAENPEIAKEVSENCENTLTAKISNIESLDNVTERSEHTLDNHKSTPEMEDV 127

```
QY 293 EPESPMVDV-----DNSRNSCODSRADEETSVPVF-DEQD----- 324
Db 128 NNSKSNIDVAINDEDEDELVLBNENKEMRGEGVQVQSOLFADDDQELIYEPGIMKDTTQ 187
QY 325 -----DRSSQTANKLSSCOAREADGDLRKRYLTGSEVRLHFOFEGENNAGTSDLNAPSG 380
Db 188 LDITDSEVETAQKEMIEEADS-----TFVGEDSKATKTVRTSSSS 230
QY 381 NSSSLNVECRSKQHGR--DSKITDHPMRISKSE-----DRRKEQCEVRHOR 426
Db 231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLPDLNKVDPDNRNYRCTI- 285
QY 427 TERKIPKYPNLPPEKKWLTPIEMRMKPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486
Db 286 -----PNFPASQ-----GKLRDNRVGPXIVL----- 307
QY 487 PKPFFTHYKDLMDNKHVMPCEQNLYPVEDRNGER-----TAG-- 525
Db 308 -----PQRWREF-----DSGRRRDSYFYFKRKLGYLKYCTTGYF 344
QY 526 -----SRWELIQTALLNKTRPQNLKDALIKYNVA-----YSKKWDFALVDFWDKVL 573
Db 345 MFVGLLHNNWEF-----DPDITYKLPALEMYIKEMSELVGRVEVLEKFA 388
QY 574 EEAEAGHLYQSILPDMVKIALCLPNICTQPIPLLKQKNHVSMTMSQEQIASILANAFECT 633
Db 389 RVARIKTAEDILPE--RIYRLVGDV-----ESATLSHKQCAALVARMFF-- 431
QY 634 FPRNNAKMSYSSYPDINFNRLFEGRSRKPPEKLTLCYFRVRVTEKPTGLVTF--TR 691
Db 432 -----ARPDSPFS-----FCRILSSDKSICVEKLFKLTFTYDKMSMDPPDGAVSFRLT 480
QY 692 QSLDEF-PEW-ERCEKPLTRLHVTEGTTIEGNGRGMLOVDPANRFVGGVGTGAGLVQBEI 749
Db 481 MDKDTFNEEMWKKKLSLPEVEFDEMLIEDTAL-CTQVDPANHLGGVGLNHSVQBEI 539
QY 804 DMQORCTEIVADALHFR--YLDQFVPEKVRRELNKAYCGFLRPGVSENLASA 858
Db 600 DRFGRLRVETIADAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQCPKFTNI-PIV 658
QY 859 TGNWCGAGFGGADARLKALIQILAAAAERDVVYTFGDSLMRDIYSMHTFLTERKLDVG 918
Db 659 TGNWCGAGFGGADARLKALIQILAAAAERDVVYTFGDSLMRDIYSMHTFLTERKLDVG 918
QY 919 K 919
Db 719 K 719
```

RESULT 15  
US-09-511-507-10  
; Sequence 10, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10

```
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-10  
  
Query Match  
Best Local Similarity 21.9%; Pred No. 5.2e-34; Length 726;  
Matches 184; Conservative 124; Mismatches 278; Indels 255; Gaps 32;  
  
QY 199 EBDNRNQOFLTPIKLANTKPTVGDQARSNCOSRSVKDCTCQCEVEDVLPESPLS 258  
Db 14 QDEKDYEDYV-GVGFHQVPTM---KRRKLTGHTNTTESKEDPE--EPKSRDVFVSSQSS 67  
QY 259 DVGADIDGTGPX-----NDNKLTQOESSLGD---SPPEKES 292  
Db 68 DESQSDSAENPEIAKEVSENCENLTETLKISNIESLDNVTSEHTLDNHKSTPEMEDV 127  
QY 293 EPESPMVDV-----DNSRNSCODSRADEETSVPVF-DEQD----- 324  
Db 128 NNSKSNIDVAINDEDEDELVLBNENKEMRGEGVQVQSOLFADDDQELIYEPGIMKDTTQ 187  
QY 325 -----DRSSQTANKLSSCOAREADGDLRKRYLTGSEVRLHFOFEGENNAGTSDLNAPSG 380  
Db 188 LDITDSEVETAQKEMIEEADS-----TFVGEDSKATKTVRTSSSS 230  
QY 381 NSSSLNVECRSKQHGR--DSKITDHPMRISKSE-----DRRKEQCEVRHOR 426  
Db 231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTEGNLTLPDLNKVDPDNRNYRCTI- 285  
QY 427 TERKIPKYPNLPPEKKWLTPIEMRMKPRCGIHLPSLRPSASHTVTVRVDLLRAGEV 486  
Db 286 -----PNFPASQ-----GKLRDNRVGPXIVL----- 307  
QY 487 PKPFFTHYKDLMDNKHVMPCEQNLYPVEDRNGER-----TAG-- 525  
Db 308 -----PQRWREF-----DSGRRRDSYFYFKRKLGYLKYCTTGYF 344  
QY 526 -----SRWELIQTALLNKTRPQNLKDALIKYNVA-----YSKKWDFALVDFWDKVL 573  
Db 345 MFVGLLHNNWEF-----DPDITYKLPALEMYIKEMSELVGRVEVLEKFA 388  
QY 574 EEAEAGHLYQSILPDMVKIALCLPNICTQPIPLLKQKNHVSMTMSQEQIASILANAFECT 633  
Db 389 RVARIKTAEDILPE--RIYRLVGDV-----ESATLSHKQCAALVARMFF-- 431  
QY 634 FPRNNAKMSYSSYPDINFNRLFEGRSRKPPEKLTLCYFRVRVTEKPTGLVTF--TR 691  
Db 432 -----ARPDSPFS-----FCRILSSDKSICVEKLFKLTFTYDKMSMDPPDGAVSFRLT 480  
QY 692 QSLDEF-PEW-ERCEKPLTRLHVTEGTTIEGNGRGMLOVDPANRFVGGVGTGAGLVQBEI 749  
Db 481 MDKDTFNEEMWKKKLSLPEVEFDEMLIEDTAL-CTQVDPANHLGGVGLNHSVQBEI 539  
QY 804 DMQORCTEIVADALHFR--YLDQFVPEKVRRELNKAYCGFLRPGVSENLASA 858  
Db 600 DRFGRLRVETIADAILFKGSKLDCQTEQLNKANIIREMKKASIGFMSQCPKFTNI-PIV 658  
QY 859 TGNWCGAGFGGADARLKALIQILAAAAERDVVYTFGDSLMRDIYSMHTFLTERKLDVG 918  
Db 659 TGNWCGAGFGGADARLKALIQILAAAAERDVVYTFGDSLMRDIYSMHTFLTERKLDVG 918  
QY 919 K 919  
Db 719 K 719
```

Search completed: May 26, 2004, 18:49:30  
Job time : 35.0181 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 80.7642 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-6  
Perfect score: 5164  
Sequence: 1 MSAGPGWECTKARWGAGT.....YHAVSSAETDMPGQKAGT 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB ID	Description
1	5164	100.0	968	9	US-09-973-451-6	Sequence 6, Appli
2	4395	85.1	976	9	US-09-973-451-4	Sequence 4, Appli
3	4332.5	83.9	977	9	US-09-973-451-2	Sequence 2, Appli
4	1074.5	20.8	768	9	US-09-973-451-8	Sequence 8, Appli
5	579	11.2	546	12	US-10-425-114-60000	Sequence 60000, A
6	443.5	8.6	726	9	US-09-973-451-10	Sequence 10, Appl
7	220.5	4.3	200	12	US-10-424-599-184988	Sequence 184988,
8	176	3.4	6642	15	US-10-369-493-5013	Sequence 5013, Ap
9	175	3.4	180	12	US-10-424-599-156445	Sequence 156445,
10	172.5	3.3	300	12	US-10-424-599-233915	Sequence 233915,
11	172	3.3	1489	15	US-10-115-482-18	Sequence 18, Appl
12	169.5	3.3	1297	12	US-10-282-122A-71692	Sequence 71692, A
13	167	3.2	331	9	US-09-973-451-11	Sequence 11, Appl
14	167	3.2	1169	12	US-10-282-122A-71179	Sequence 71179, A
15	163.5	3.2	1210	14	US-10-205-823-264	Sequence 264, App

16	161	3.1	748	9	US-09-864-761-43244	Sequence 43244, A
17	159.5	3.1	1501	9	US-09-924-154-17	Sequence 17, Appl
18	159	3.1	1462	14	US-10-287-218-17	Sequence 17, Appl
19	159	3.1	2677	15	US-10-144-194A-22	Sequence 22, Appl
20	159	3.1	2697	15	US-10-144-198-12	Sequence 12, Appl
21	158.5	3.1	4723	15	US-10-359-012-8	Sequence 8, Appli
22	156	3.0	691	15	US-10-108-260A-4345	Sequence 4345, Ap
23	150	2.9	447	10	US-09-794-422-44	Sequence 44, Appl
24	150	2.9	1115	13	US-10-153-273-2	Sequence 2, Appli
25	148	2.9	1389	12	US-10-282-122A-70151	Sequence 70151, A
26	147.5	2.9	1358	15	US-10-369-493-1658	Sequence 1658, Ap
27	147.5	2.9	1973	16	US-10-197-824-34	Sequence 34, Appl
28	147.5	2.9	2047	12	US-10-346-863-2	Sequence 2, Appli
29	147.5	2.9	3863	16	US-10-197-824-7	Sequence 7, Appli
30	145.5	2.8	1881	14	US-10-032-585-7646	Sequence 7646, Ap
31	144.5	2.8	747	12	US-10-425-114-67803	Sequence 67803, A
32	144	2.8	26	9	US-09-973-451-21	Sequence 21, Appl
33	144	2.8	29	9	US-09-973-451-12	Sequence 12, Appl
34	143	2.8	1253	14	US-10-363-798-2	Sequence 2, Appli
35	143	2.8	2158	15	US-10-341-434-222	Sequence 222, App
36	143	2.8	2158	15	US-10-144-198-34	Sequence 34, Appl
37	143	2.8	2265	15	US-10-144-198-35	Sequence 35, Appl
38	142	2.7	1237	15	US-10-108-260A-3795	Sequence 3795, Ap
39	141.5	2.7	677	14	US-10-060-036-180	Sequence 180, App
40	140.5	2.7	3067	10	US-09-949-029-18	Sequence 18, Appl
41	140	2.7	1309	15	US-10-369-493-1656	Sequence 1656, Ap
42	140	2.7	1325	9	US-09-864-761-35612	Sequence 35612, A
43	140	2.7	1460	15	US-10-295-027-428	Sequence 428, App
44	139.5	2.7	498	12	US-10-424-599-196154	Sequence 196154,
45	139.5	2.7	791	15	US-10-144-194A-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1

US-09-973-451-6  
; Sequence 6, Application US/09973451  
; Patent No. US2002013228A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 6  
; LENGTH: 968  
; TYPE: PRT  
; ORGANISM: Mus musculus  
; FEATURE:  
US-09-973-451-6

Query Match	100.0%;	Score 5164;	DB 9;	Length 968;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 968;	Conservative	0;	Mismatches	0;
			Indels	0;
Qy	1	MSAGPGWECTKARWGAGTSAFTASDSRSPGQRRLDPKDPVQFRVPPSPACVSG	60	
Db	1	MSAGPGWECTKARWGAGTSAFTASDSRSPGQRRLDPKDPVQFRVPPSPACVSG	60	
Qy	61	RAGPHRGNTSFVFKQTKTTTWMDTKGTABESKENNTRIDSMSSVKQNFPHKV	120	

Db 61 RAGPHRGNATSPVFKQKTIITWMDTKGPKTAESKENNTRIDSMSSVQKDNFYPHKV 120  
QY 121 EKLENVPOINLDKSTKSSQYLNQOQTASVCKWQNEKGHAEQIOLASPPAGTLPKQLS 180  
Db 121 EKLENVPOINLDKSTKSSQYLNQOQTASVCKWQNEKGHAEQIOLASPPAGTLPKQLS 180  
QY 181 NANIGQSPTDHDSDTDEEDRDNOQFLTPIKLANTKPTVGDQARSCKSGSRQSVKD 240  
Db 181 NANIGQSPTDHDSDTDEEDRDNOQFLTPIKLANTKPTVGDQARSCKSGSRQSVKD 240  
QY 241 CTGCOQEEVDVLPESPLSDVGAEDITGPKNDKNTGQESSIGDSPPEKSESPSPMDV 300  
Db 241 CTGCOQEEVDVLPESPLSDVGAEDITGPKNDKNTGQESSIGDSPPEKSESPSPMDV 300  
QY 301 DNRNSCODSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLTGKSEVRHL 360  
Db 301 DNRNSCODSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLTGKSEVRHL 360  
QY 361 FOFEGENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFMRISKSEDRKQEC 420  
Db 361 FOFEGENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFMRISKSEDRKQEC 420  
QY 421 EYRHQTERKIPKYIPPNLPPEKKWLGTPIEMRKMPRCGIHLPSLRPSASHTVVRDL 480  
Db 421 EYRHQTERKIPKYIPPNLPPEKKWLGTPIEMRKMPRCGIHLPSLRPSASHTVVRDL 480  
QY 481 LRAGEVPKPPHYKDLWNKHVQKPCSEONLYPVEDENGERTAGSRWELIQTALINKPT 540  
Db 481 LRAGEVPKPPHYKDLWNKHVQKPCSEONLYPVEDENGERTAGSRWELIQTALINKPT 540  
QY 541 RPQNLKDALIKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVKIALCLPNIC 600  
Db 541 RPQNLKDALIKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVKIALCLPNIC 600  
QY 601 TQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
Db 601 TQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLFEGR 660  
QY 661 SRKPEKLTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELHVTYEGTIEG 720  
Db 661 SRKPEKLTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELHVTYEGTIEG 720  
QY 721 NGRGMLQVDFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNECLIIITGTEQ 780  
Db 721 NGRGMLQVDFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNECLIIITGTEQ 780  
QY 781 YSEYTGATYRWARSHEDSGSKDQWRCTEIVADLHFRYLDQFVPEKVRRELNKA 840  
Db 781 YSEYTGATYRWARSHEDSGSKDQWRCTEIVADLHFRYLDQFVPEKVRRELNKA 840  
QY 841 YGFLRPGVPSNLAVATGNCGAGFGDARLKALIQILAAAAERDVVYFTFGDSELM 900  
Db 841 YGFLRPGVPSNLAVATGNCGAGFGDARLKALIQILAAAAERDVVYFTFGDSELM 900  
QY 901 RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECNCSTPGPDIKLYPIYHAVESSAETTD 960  
Db 901 RDIYSMHTFLTERKLDVGKVKYLLRLRYNEECNCSTPGPDIKLYPIYHAVESSAETTD 960  
QY 961 MPQQAAGT 968  
Db 961 MPQQAAGT 968

## RESULT 2

US-09-973-451-4  
; Sequence 4, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES,  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-973-451-4

Query Match 85.1%; Score 4395; DB 9; Length 976;

Best Local Similarity 85.7%; Pred. No. 0;

Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MSAGRGWPECTKA-RWGAAGTSAPTASDSRSPGQREVLDPKQAPVQFRVPPSPACV 59

Db 1 MNAGGCEPCTKATRWGAA-TTSPAASDARSFGRQRRVLPKQAHVQFRVPPSPACV 59

QY 60 GRAGPHRGNATSPVFKQKTIITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116

Db 60 GRAGPHRGNATSPVFKQKTIITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116

QY 60 GRAGPHRGNATSPVFKQKTIITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116

Db 60 GRAGPHRGNATSPVFKQKTIITWMDTKGPKTAESE---SKENNNTRIDSMSSVQKDNFY 116

QY 117 PHKEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGHAEQIOLASPPAGTLP 176

Db 117 PHKEKLENVQNLNLDKSPTEKSSQYLNQOQTASVCKWQNEKGHAEQIOLASPPAGTLP 176

QY 120 QHNEKLVNVQSLSLDKSLTEKSTQYLNQOQTAAQCKWQNEKGHAEQIOLASPPAGTLP 179

Db 120 QHNEKLVNVQSLSLDKSLTEKSTQYLNQOQTAAQCKWQNEKGHAEQIOLASPPAGTLP 179

QY 177 KQLSNANIGQSPTDHDSDTDEEDRDNOQFLTPIKLANTKPTVGDQAR---SNCKCSG 233

Db 177 KQLSNANIGQSPTDHDSDTDEEDRDNOQFLTPIKLANTKPTVGDQAR---SNCKCSG 233

QY 180 EQFSNANIDRSPQNDHSDTDEEDRDNOQFLTPIKLANTKPTVGDQAR---SNCKCSG 239

Db 180 EQFSNANIDRSPQNDHSDTDEEDRDNOQFLTPIKLANTKPTVGDQAR---SNCKCSG 239

QY 234 SRQSVKDCGTCQOQEEVDVLPESPLSDVGAEDITGPKNDKNTGQESSIGDSPPEKSE 293

Db 234 SRQSVKDCGTCQOQEEVDVLPESPLSDVGAEDITGPKNDKNTGQESSIGDSPPEKSE 293

QY 240 SCHPQEDCASQOQDEIDVVPKSPLDVQSEEDVGTGSKNDKNTGQESSIGDSPPEKSE 299

Db 240 SCHPQEDCASQOQDEIDVVPKSPLDVQSEEDVGTGSKNDKNTGQESSIGDSPPEKSE 299

QY 294 PESPMVDVNSRNSQDSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLT 352

Db 294 PESPMVDVNSRNSQDSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLT 352

QY 300 PESPMVDVNSRNSQDSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLT 359

Db 300 PESPMVDVNSRNSQDSEADETSPVDFODDRSSQTANKLSSCOAREADGDLKRYLT 359

QY 333 KGEVRLHFOPE-GENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFWRISK 411

Db 333 KGEVRLHFOPE-GENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFWRISK 411

QY 360 KGEVRLHFOPE-GENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFWRISK 419

Db 360 KGEVRLHFOPE-GENNAGTSDLNAPSGNSSLNVECRSKQHGKRDSKIITDHFWRISK 419

QY 412 SEDREKQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRMPCGIRLPLRPSAN 471

Db 412 SEDREKQCEVRHQRTERKIPKYIPPNLPPEKKWLGTPIEEMRMPCGIRLPLRPSAN 471

QY 420 AEDRKEQWEVTEKPTGLVTTROSLEDPPMERCEKPLTELHVTYEGTIEG 531

Db 420 AEDRKEQWEVTEKPTGLVTTROSLEDPPMERCEKPLTELHVTYEGTIEG 531

QY 472 HTVTVRVLLRAGEVPKPPHYKDLWNKHVQKPCSEONLYPVEDENGERTAGSRWELI 539

Db 472 HTVTVRVLLRAGEVPKPPHYKDLWNKHVQKPCSEONLYPVEDENGERTAGSRWELI 539

QY 532 QPALLNKTRPONLKDAILKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVK 591

Db 532 QPALLNKTRPONLKDAILKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVK 591

QY 540 QPALLNKTRPONLKDAILKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVK 599

Db 540 QPALLNKTRPONLKDAILKYNVAYSKWDFALVDFWDKVLEAEQAHLQSIILPDMVK 599

QY 592 IALCLPNICTQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651

Db 592 IALCLPNICTQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651

QY 600 IALCLPNICTQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659

Db 600 IALCLPNICTQPIPLKQKQNMHVSVMSSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659

QY 652 NFNRLFEGRSRKPKLKTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELH 711

Db 652 NFNRLFEGRSRKPKLKTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELH 711

QY 660 NFNRLFEGRSRKPKLKTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELH 719

Db 660 NFNRLFEGRSRKPKLKTLCYFRFRVTEKPTGLVTTROSLEDPPMERCEKPLTELH 719

QY 712 VTYEGTIEGNRGMLOVDGFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNE 771

Db 712 VTYEGTIEGNRGMLOVDGFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNE 771

QY 720 VTYEGTIEGNRGMLOVDGFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNE 779

Db 720 VTYEGTIEGNRGMLOVDGFANRFGVGGVTGAGLVQEBIRFLINPELIVSRFLTEVLDHNE 779



QY 772 CLITGTQYSEYTGATYRWASHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 831  
 Db 780 CLITGTQYSEYTGATYRWASHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 839  
 QY 832 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVY 891  
 Db 840 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVY 899  
 QY 892 FTFGSELMDIYSMHITLTERKLDVGKVKYKLLRLRYNEECNCSTPGDILKLPFIYHA 951  
 Db 900 FTFGSELMDIYSMHITLTERKLDVGKVKYKLLRLRYNEECNCSTPGDILKLPFIYHA 959  
 QY 952 VESSAETDMPGQKAGT 968  
 Db 960 VESSAETADHSQRTGT 976  
 RESULT 3  
 US-09-973-451-2  
 ; Sequence 2, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-973-451-2  
 Query Match 83.9%; Score 4332.5; DB 9; Length 977;  
 Best Local Similarity 84.0%; Pred. No. 0;  
 Matches 820; Conservative 57; Mismatches 90; Indels 9; Gaps 5;  
 QY 1 MSAGPGWEPTCK-ARWGAAGTSAPTASDSRSPGRRVLDPKADPQVFRVPPSPACVS 59  
 Db 1 MSAGPGCEPCTKRRWDAAATSPPAASDARSFGRQRRVLDKADPQVFRVPPSSGCA 60  
 QY 60 GRAGPHRGATSFVFQKTIITMDTKGPXTAESE---SKENNTRIDSMSSVQKDNFY 116  
 Db 61 GRAGQHRGATSILVFQKTIITMDTKGIVTESILHSKNNNTREESNMSSVQKDNFY 120  
 QY 117 PHKVEKLENVQNLNPKSPTEKSSQVILNQOATASVCKWQNEGHAEOLLASEPPACTPLP 176  
 Db 121 QHNMEKLENVQNLNPKSPTEKSSQVILNQOATASVCKWQNEGHAEOLLASEPPACTPLP 180  
 QY 177 KOLSNANIGOSPTDHDSDTDEEDRNOQFLTPILKANTKPTVGGQ---ARSNCKCSG 233  
 Db 181 EQFSNVNDGSSPKDDHSDTNSSESDNQFLTHVLKLANAKQTMEDQGREASHQCKG 240  
 QY 234 SRQSVKDCQCEQREVDVLPESLSDVGAEDIGTPKNDKLTGQSSSLGDSPPFEKESE 293  
 Db 241 ACPAEACACQCEQREVDVLPESLSDVGAEDIGTPKNDKLTGQSSSLGDSPPFEKESE 300  
 QY 294 PESPMVDNRSNCSQSEADEETSVPFDEQDRSS-OTANKLSQCAAREADGLRKLYLT 352  
 Db 301 PESPMVDNRSNCSQSEADEETSVPFDEQDRSS-OTANKLPKRFQREADTLRKSSA 360

QY 353 KGSEVRLHFOPE-GENNAGTSDLNNAKPSGSSSLNVECRSSKQHGKRDskLTDHFMRIK 411  
 Db 361 KGSEVRLHFOPEGGESRAGMNDVNAKPGSTSSLNVECRSSKQHGKRDskLTDHFMRIK 420  
 QY 412 SEDRRKQCEVRHQRTERKIPKYPNLPKPKWLGTPIEEMRMKPRCGIHLPSLRPSAS 471  
 Db 421 AEDRRKQCEVRHQRTERKIPKYPNLPKPKWLGTPIEEMRMKPRCGIHLPSLRPSAN 480  
 QY 472 HTVTVRVLDLRAEVPKPPFTHYKDLMDNKHVKMPKPCSEONLYPVEDENGERTAGSRWELI 531  
 Db 481 HTVTIRVLDLRAEVPKPPFTHYKDLMDNKHVKMPKPCSEONLYPVEDENGERTAGSRWELI 540  
 QY 532 QTALLNKTRPQNLKDALIKNVAYSKWDALVDKWDKVLLEAARAOHLQSLPDMVK 591  
 Db 541 QTALLNRLTRPQNLKDALIKNVAYSKWDALVDKWDKVLLEAARAOHLQSLPDMVK 600  
 QY 592 IALCLPNICTOPIPLLKQKMHVSITMSQEQIASLLANAFCTFFPRNNAKMSSESYSDI 651  
 Db 601 IALCLPNICTOPIPLLKQKMHVSITMSQEQIASLLANAFCTFFPRNNAKMSSESYSDI 660  
 QY 652 NFNLFEGRSSRKPKLTKLFCYFARVTEKPTGLVTTROSLEDFFPWERCEKLTILH 711  
 Db 661 NFNLFEGRSSRKPKLTKLFCYFARVTEKPTGLVTTROSLEDFFPWERCEKLTILH 720  
 QY 712 VTTEGTIEGNGRGMLOVDPANRFVGGVGTAGLVQEEIRFLINPELIYSRLFTVLHDNE 771  
 Db 721 VTTEGTIEGNGRGMLOVDPANRFVGGVGTAGLVQEEIRFLINPELIYSRLFTVLHDNE 780  
 QY 772 CLITGTQYSEYTGATYRWASHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 831  
 Db 781 CLITGTQYSEYTGATYRWASHEDSGKDDWQRCTEIVADALHFRYLDQFVPE 840  
 QY 832 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVY 891  
 Db 841 KVRRELKAYCGFLRPGVPSNLSAVATGNWCGAFGGDARLKALIQILAAAAERDVVY 900  
 QY 892 FTFGSELMDIYSMHITLTERKLDVGKVKYKLLRLRYNEECNCSTPGDILKLPFIYHA 951  
 Db 901 FTFGSELMDIYSMHITLTERKLDVGKVKYKLLRLRYNEECNCSTPGDILKLPFIYHA 960  
 QY 952 VESSAETDMPGQKAG 967  
 Db 961 VESCTQTINQPGRTG 976  
 RESULT 4  
 US-09-973-451-8  
 ; Sequence 8, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-973-451-8

Matches	150;	Conservative	69;	Mismatches	154;	Indels	70;	Gaps	12;
545	LKDAILKNVAYS-----	KKWDFETALDVFWDKVLLEAEAOHLQCSILPDMVKIALCLPN	598						
88	LADALTDLRLALAPLPPRAADGLAL--FFDDLLSRAAARGWFSEVVVNLRLLRLPT	145							
599	ICTQPIPLLKQXNNHS-----	VTMSGEQIASLLANAFCTCFPRNAKMKE	644						
146	LLEDHY-----	AKAGHGASGLRYMASQDAGVLLSQELVAALLTCALFCFLPTAGRAQ---	198						
645	YSSVPDINENRNF----	EGRSRRPEKLKTLCFCVFRVTEKKPTGLVTFPROSL-----	694						
199	-ACLPITNFDGFLFAPLIHNRASQB-QKVRCVLHVFERVTSMTGLVSFKVLPRRAL	256							
695	-----EDFPWERCEKPLTRLHVITYEGTIEGNRGMLQVDFAHFVGCGVTGAGLVQEE	748							
257	SDGVYPDPIDHAMVASSAPLCQFTVSSGFIEDQEQAQLQVDFANKLGGGALSVCVQEE	316							
749	IRFLINPELIVSLPFTEVLDRNECLIITTEGYSEYTYGAASYTYRWARSHEGSEKDDWQR	808							
317	IREFMINPELTVMGLFISCMEDNEATIEFGAERFSQYMVGYSFRPVGDYLDTKPFDSMGR	376							
809	RCTEIVAIIDALHFERRLDQFVPEKYVRRELNKAYCCFLR-----	PG 848							
377	RRTRIIVAIDALDCPAKL-HYESGCLLRNVNKAFCFFQPSQKHLYAKLFQDLHNKDDFSS	435							
849	VPSENLSAVATGNWGCGAGDGDAIKALKIQLIAAAAAERDVV-YPTFGSELMRDIYSMH	907							
436	INSSEYGVSTGNWCGAGFNPEIKSMIQWIAAQALRPFNYYNYTFEDVSIQR-LEEVI	494							
908	TFLTERKLDVGKVKYKILLRYNE	930							
495	QWIRLHCWTVGELWHLMLEYSSQ	517							
<b>RESULT 6</b>									
US-09-973-451-10									
; Sequence 10, Application US/09973451									
; Patent No. US2002012328A1									
; GENERAL INFORMATION:									
; APPLICANT: JACOBSON, Myron K.									
; APPLICANT: JACOBSON, Elaine L.									
; APPLICANT: AM, Jean-Christophe									
; APPLICANT: LIN, Winston									
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE									
; TITLE OF INVENTION: (PARG) ENZYMES,									
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE TO SAME									
; FILE REFERENCE: N/A 201									
; CURRENT APPLICATION NUMBER: US/09/973,451									
; CURRENT FILING DATE: 2001-10-09									
; PRIOR APPLICATION NUMBER: US/09/302,812									
; PRIOR FILING DATE: 1999-04-30									
; PRIOR APPLICATION NUMBER: 60/083,768									
; PRIOR FILING DATE: 1998-05-01									
; NUMBER OF SEQ ID NOS: 38									
; SEQ ID NO 10									
; LENGTH: 726									
; TYPE: PRT									
; ORGANISM: Caenorhabditis elegans									
; FEATURE:									
US-09-973-451-10									
Query Match                      8.6%; Score 443.5; DB 9; Length 726;									
Best Local Similarity        21.9%; Pred. No. 3.1e-26;									
Matches 184; Conservative     124; Mismatches 278; Indels 255; Gaps 32									
QY	199	EBDRNQOFLTPKIANKTKTVGDGQARSNCKSGRSQSVKDTGCGQQEQEVVLPSPUS	258						
DB	14	QDEKDYEDY-VGVFAHQVPTM---KRRKLTGHNTGESKDOPE--EPKSRUVFVSSQSS	67						
QY	259	DVGREDITGTGPK-----NDNKLKTQESSLGD---SPPFKES	292						





```

; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/285,325
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: 60/285,890
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286,068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286,292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/287,213
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: 60/288,257
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/291,134
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/291,725
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/294,771
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/296,965
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: 60/299,128
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 149
; SEQ ID NO 18
; LENGTH: 1489
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-115-482-18

Query Match      3.3%; Score 172; DB 15; Length 1489;
Best Local Similarity 19.5%; Pred. No. 0.00049;
Matches 161; Conservative 115; Mismatches 301; Indels 248; Gaps 39;

QY 75 KQKTIITWM---DTKGPKTAESKENNTRIDSMSSVQKDNF-YPHKVEKLENVPLN 130
DB 418 KQRLVNTYMPQPRGDEQSDSDSFNPLTWEEQRQORMTVAPEFEDKEDDENAGKVK 477
QY 131 LD-KSPTEKSSQVYNQOQ---TASVC-KWQNEGHAEQL-----LASEP 169
DB 478 LSCQAPWERGQGITLQPARLSGDCCTFWARCDQIQDMPVQNDPQLAWGCISGLQOER 537
QY 170 PACTPLPKQLSNA-----NIGQSPHTDDHSDTDHEEDRNQQLTPIKLANTK 217
DB 538 SMCPTPLVAAQSTTLPLSLGRQVEINKRYP-----TYPEDLKN--WKSVQNLVVK 588
QY 218 PTVGDGQARNCCKSGRSQVQKCTGQOQBEVDVLPESPLSDVGAEDIGTPKNDKLTG 277
DB 589 PSHGVRVNSN-PTANTQITVKE---KYEHWKVPAPK---EITVEDSFVHPANEMRIGE 640
QY 278 QESSLGDSPPF-----EKSEPEPMVDVNSRNSQDSADETSPTVDFEQDDRS 327
DB 641 LHPSLAETPLYPKLLVLLGDKKESSTDESEYDKTHCLNNSVSSCTYSYDSP--SQASSGS 698
QY 328 SQTANKLSSCOA-----REA--DGLRKR----- 349
DB 699 SNTRVKVGLQTAKDAVHNSLWGNRIAPSPQPLDSKPLLSQREAVPPGNIPORPLRP 758
QY 350 -----YLTGSEVRHLHFQEGENNAG--TSDLNAPSGNSSSLINVECRSSKHQGRDS 400
DB 759 MSDTFTDNWTDG-----HY-----DNTGFAVEETTAENANSNPLSSKRSRSTSSHGRLP 809
QY 401 KITDFMRISKSDRRKQCEVRHQRTKIPKYPNPLPPEKKWLTG--PIEMRKNMPP 459

```

```

DB 810 IRQDRIVGVPLELE-----OSTHRHTPETEVPSPNPWQNTWTRTSPSPFFEDRTAFPS 859
QY 459 CGIHLPSLRPSASHTVTVRVLDLLRAGEVKPF-----PTHYKDLWDNKHVKMPCSEQNLYP 514
DB 860 KLETTPTTSPLPERKEHIK-----ESTEIFSPSPGVPWEYHDSNPNRSLNSVFSQIHCRP 915
QY 515 VEDENGERTAGSRWE-----LIOTALLNKTRPNQNLKDAL---KYNVA---YSKKWDF 563
DB 916 -BSSKGVISIKSTERLSPLMKDIKSNKFKKSQSIDIDITGYTKVYNIPLNENYASGSDHL 974
QY 564 ALVDFWPKVLEAEAOHLYQSILPDMVKIALCLPNICTOPIPLLKOKM--NHSVITMSQEQ 621
DB 975 GSHERPDKML--GPEHGMSSMSR-----SOSVPMDDDEMLTYGSSKGPQQ 1018
QY 622 IASLANA--FFCTF--PRNNAKMKSE-----YSSY 648
DB 1019 KASMTKKVYQFOSFNPOGSVEVKAERIPPPFQHNPEYVQQAQKNIADLISPRARGY 1078
QY 649 P-----DINFNRLFGRSS-----RPEKL--KTLFCYFRRVTTEKPTGL 686
DB 1079 PPMEQMFSPSPSVNEDAVVNAQFASQAGARGLREADSLVSATAMAFRRVNEP----- 1133
QY 687 VTFTRQSLDFEWEKCEKPLTE-----LHVTYEGTIEGNGR 723
DB 1134 -----HELPTDRIYGRPPYRGGLDROSSVTVITVESQFLKXNGR 1170

RESULT 12
US-10-282-122A-71692
; Sequence 71692, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71692
; LENGTH: 1297

```

```

Db      1094 ELADLMMAAPQVVEGSIARIAQKARACGHMLVATQPSVNVITGLIKSNITPIAFMVS 1153
      859 TWGCGAGFGDGRDLKALQILAAAAA-----DVVYTFGDSLEMRDIYSNMHTPLTER 913
      1154 S-----SVDSR-----TILDSGAERLLGVGDMLYLGSG---MNKPIRVQGTFSDD 1197

QY      914 KLD 916
      1198 EID 1200

Db

RESULT 13
US-09-973-451-11
; Sequence 11, Application US/09373451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY(ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTI
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 11
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; US-09-973-451-11

Query Match 3.2%; Score 167; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      762 LFTFVLHDNECLITGTQYSEYTGVAETVR 792
      1 LFTFVLHDNECLITGTQYSEYTGVAETVR 31

Db

RESULT 14
US-10-282-122A-71179
; Sequence 71179, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

```

```

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71179
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-71179

Query Match
  3.2%; Score 167; DB 12; Length 1169;
Best Local Similarity 19.1%; Pred. No. 0.00084;
Matches 182; Conservative 128; Mismatches 290; Indels 352; Gaps 44;

Qy 90 TAESSEKNNTRIDSMSSVQK-----DNFYPHKVEKLENPQNLDKSPKTSQY 142
Db 348 TGESSLLENESQDSSNSLEKQSSNIDNKEAQNNTPLNFYEEIDLQ---TTSVYK 404

Qy 143 LNOOQTASVCKWQNGKHAQELLASEPPAGTPLPKQLSNANIGOSPHYDHSYD---H 198
Db 405 VNEBET-----ESKNDLVS-----NHYHSNDDAVEDAEYH 438

Qy 199 EEBEDNOQFLTPIKLANTK-PTVGDGQARSCKSGSRQSV-----KCTGCGQE 247
Db 439 ELDDNRQO-----NOSNQDDIISKSTSMYDANASVDNNTERRAKSNEDKNDTEIT 493

Qy 248 EVDVLPESPLSDVGAEE-----DGTGPKNDKL-----TCQ--ESSLGDSPFF-- 288
Db 494 HLDGTTSAKVSDEKTESNTNHLQDKNVKLVKXVNSLSSNSDGTQKQKQFGSRPFV 553

Qy 289 -----EKSEBESPMVDVNSRNSQDSEAD-----EETSVPFD 321
Db 554 LMTFSDKKRMDQNHKKVSVFELPKPKQANANHRKDSKNSKSEBFKQINTNRETDSNYS 613

Qy 322 E---QDSSOTANKLSSQAREADGDLRKRYLTKGSEVRHLHFOGEGNNAGTSDLNAP 378
Db 614 SNGLEHDMNSSDHEVYETPSKQDEQIQK-----LQDDFFENANAKIN--NSNE 663

Qy 379 SGNSSSLNVECRSKQHGKRSKTDHFMRTKSEDRKKEQCEVHRHQRTERKIPYIPPN 438
Db 664 TGNQSNIS-----HSKRQSYSTNE-----SKNIDTQTSNSTSNQNFQ----- 702

Qy 439 LPPEKKWLGTPIEWKQMPRCGHLPSLRPSASHVTIVRVLLRAGEVPKPPFTHYKDL- 497
Db 703 -----IRKGP--NKLPSYQ-----LL---EAPEP---HEKDDQ 728

Qy 498 W-DNKHVMQPCSEONLYPVEDENGERTAGSRWELIQTALLNKFTRPQMLKADAILKYNVAY 556
Db 729 WIDNK-----KQELNDALYFNV-- 745

Qy 557 SKKWDFTALVDFWVKVLEEAQAHLQYQILPDMVKIALCLPNICTQPTPLLKQKNHVS 616
Db 747 -----PAEVKNVTEG-----PSVTRFELSVEKGVKVRKIT 776

Qy 617 MSQFIASLLANAFCTFPRRNAMKSEYSSYPDINFNRLPEGSSRKPPEKLTFLCYFR 676
Db 777 ALODDIKMAA-----AK-----DIRTEAPIPGTSLVGIE----- 806

```

```

Qy 677 RVTEKKPTGLVTFTRQSLDFPEWERCEKPLTRLHVTVYEGTIEGNGMLQVDFANRFVG 736
Db 807 -VPNQNTKV---NLSIIESPKNVTSKLT-----VAMGYRINNEPLMDIAKTPHALI 858

Qy 737 GGVGTAGLV-----QBEIR-FLINPELI-----VSRLEFTEVLDH 769
Db 859 AGATGCKSVCSINSILMSILYKKNHPELRLLLIDPKMVELAPYNDLPHLVSPVITDKAA 918

Qy 770 NECLITITGTYSEYTGATYRWARSHEDGSEKDDQWRCTEIV-AIDAL-HFERYLDQ 827
Db 919 TQSLKWAVEEMEKRYKLFQYH--VRNITAFNKAPYEQRMKPIVIVIDEADLMMAPQ 976

Qy 828 FVPEKVRRELKNA-----YCGFLRPGVPSENLAVATGNWCGAFGG 869
Db 977 DVEQSIAIAQKARACGIEHMLVATQPSVNVITGLIKANIPTRIAFVSS-----SV 1028

Qy 870 DARKALIQILAAAAER-----DVYFTFGDSELMEDDIYSMHFTLTERKLD 916
Db 1029 DSR-----TILDSGABRLLYGDMLYLGSQ---MNKPIRVQGTFFVSDDEID 1072

RESULT 15
US-10-205-823-264
; Sequence 264, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gamavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wonsley, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MEI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264
; LENGTH: 1210
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-205-823-264

Query Match
  3.2%; Score 163.5; DB 14; Length 1210;
Best Local Similarity 19.7%; Pred. No. 0.0017;
Matches 197; Conservative 112; Mismatches 368; Indels 321; Gaps 42;

Qy 35 QRRVLDPKDPAPVFRVPPSPSPACVSGRAGHRGNATSVFVKQKITTITMTDKPKTAESE 94
Db 260 QDSLVAPAQPPSQTFPPPSLP-----SKSVAMQKPTAYVRPMDGQDQAPSE 306

Qy 95 SKENNTRIDSMSSVQKINFYHVKLENPQNLDKSPKTSQYLNQOQTASVCKW 154
Db 307 SPELKLPLPYRQOTFEKTDL---KVPKAKLTKLRMPSQSVTEQ-----Y 349

```

QY	155	QNEGKHAEQLL-----ASEPPAGTPLP-----	176
Db	350	SNEVHCVEEILKEMTHSWPPLTAIHTPSTAEPSKFPFTKDSQHVSVTONQKOYDTSS	409
QY	177	KOLSNANTIGOSPHTDHDSDRDNCOQFLTPIK--LANTKPTV-----G	221
Db	410	KTHSNSQQTSMLEDDQLQSDSESDSEQ--TPEKPPSSSAPPAPQSLPEPVASAHSS	467
QY	222	DGOASNCKSCGSROSVDCTGCOQEEVDVL-----PESPLSDVGAEDIGTPKNDKNL	275
Db	468	SAESESSTDSDDSESSSSDSENEPLETPAPEPEPPTTNKM-----QLDNWL	519
QY	276	TQOESLGDSPPEFEKESPEPMVDVNS-----RNSCQDSEADETSPVFDEQDRSSOT	330
Db	520	T-----KVSQPAAPPEGPSTEPFRHPESKGSDDSATSQEHSESKDPPPKS	566
QY	331	ANKL-----SSQAREADGLRKRVLTKGSEVRLHFPQEGENNAGTSDLNAPK	378
Db	567	SSKAPRAPPEAPHPGKRSCKQSPAQQEPPQR-----QTVGTKQ--PKXP	608
QY	379	SGNSSSLNVECRSSKHGRDSDKITDHFMRISKSEDRKEQCEVRHQRTERRKIPKYPN	438
Db	609	V--KASARAGSRTSLQ--GEREPCLLPYGSRDQTSKDKPKVTKGRPRAAASNEPK---	662
QY	439	LPPEKKMLGTPIEMKMPRCGIHLPSLRPSASHVTVRVDLLRAGEVPKPPPTHYKDLW	498
Db	663	VPPSSE-----KKGHKSSLPAPS-----KALSGPEPAKONVEDRT	697
QY	499	DNKHVMPQCEQNLXVEDENGERTAGSRWELIOTALLNKFTRPQNLKDAILLKYNVAYSK	558
Db	698	PEHFALVPLTESQ--GPHSGSGRTSGCROAVVQEDSRKDRPLPLRDT-----	746
QY	559	KWDF TALVDFWDXLEAEAAHLYQSILPDMVKIALCLPNICTQP-----	603
Db	747	-----KLSPLRDTPPQSL-----MVKITLDLSRIQPFGKSGRQKAEKQP	791
QY	604	-----IPLLKQKM-----NHSVTM-----SQBIASLLANAFCTIF	634
Db	792	PAGKHSSSEKSSDSSSKLAKRKGAEKDCDNKIRLEKIKSQSSSSSSSHKSSKTK	851
QY	635	PRENAKMSYSSYPDINFENRLFEGSRSSRKEK--LKTILFCYFRRVTE---KKPTGLVTF	689
Db	852	PSRPSQSSKEMLP PPPVS-----SSSQKPAKPAKCR-----SRREADTCGQDPPKSASS	902
QY	690	TROSLED--FPWERCEKPLTRLHVTYEGTI-----EGNGR--GMLQVDFANR	733
Db	903	TKSNHKDSSIPKQRRVEGKSGSSSEHKSGSGDTANFPVPVSLPNGNSKFGKFPQVKFDKQ	962
QY	734	FVGGGVGTAGL--VOEERFLINPELIVSRL-----PTEVLHDNECLIIITGBOYSE---	783
Db	963	-----QADLHMREAKMKQKAEMLMTRDVGKAFKYLEAVLSFTECGIATESQSXSXA	1015
QY	784	YTGVAET-----YRWARSHEDGSEKDDWQ-----RRCTEIVADALHFRYLDQFVPE	831
Db	1016	YSVYSEIVDLIKFTMSLKSFSDATAPTQEKIFAVLCMRQCSI--LNNWAMFRCKD--IAI	1071
QY	832	KVRRELNKAYCGFLRPGVPSENTSAVATGNWGCAGFG	869
Db	1072	KYSRTLNKHF-----ESSKVAQAPSPCIASTG	1099



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:23:35 ; Search time 26.6288 Seconds

(without alignments)  
3496.721 Million cell updates/sec

Title: US-09-302-812-6

Perfect score: 5164

Sequence: 1 MSAGGWEPTKARWAGCT.....YHAVSSAETDMPQKAGT 968

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	517	10.0	997	2 B84726	probable poly(ADP-ribose)
2	443.5	8.6	726	2 T21138	hypothetical prote
3	328	6.4	364	2 A84726	probable poly(ADP-ribose)
4	176	3.4	6642	2 T29757	protein UNC-89 - C
5	170	3.3	997	2 T43523	cut17 protein - fi
6	166	3.2	1165	2 T16420	hypothetical prote
7	165	3.2	1192	2 A71623	probable secreted
8	163.5	3.2	1210	2 T39410	AP-4 protein, spli
9	163	3.2	532	2 T06029	hypothetical prote
10	161.5	3.1	2722	2 T20532	hypothetical prote
11	160	3.1	2738	2 E88320	protein F07A11.6
12	159.5	3.1	1115	2 T19137	hypothetical prote
13	159.5	3.1	1213	2 A58198	serine/proline-ric
14	159	3.1	1403	1 A47328	natural killer cel
15	157.5	3.0	2211	1 KFB05	coagulation factor
16	156.5	3.0	1001	2 T16419	hypothetical prote
17	156	3.0	1274	2 T37193	enamelin matrix pr
18	156	3.0	2526	2 T20531	hypothetical prote
19	155.5	3.0	990	2 T51618	nucleolar phosphop
20	154.5	3.0	1560	2 T42727	proliferation pote
21	153	3.0	1792	2 T08878	supervillin P205 -
22	152.5	3.0	406	2 S38170	SRP40 protein - ye
23	152	2.9	489	2 A45988	dentin matrix acid
24	152	2.9	1359	2 T34036	hypothetical prote
25	150	2.9	699	2 T38073	nucleolar phosphop
26	150	2.9	1070	2 T30848	Duffy receptor - p
27	150	2.9	1337	2 T30291	dextranase - Strept
28	149.5	2.9	2361	2 T25752	hypothetical prote
29	149	2.9	1401	2 T48079	hypothetical prote

30	149	2.9	1507	2 B47328	natural killer cel
31	148.5	2.9	2706	2 T28155	variant-specific s
32	147.5	2.9	1358	2 A29360	SIR4 protein - Yea
33	147.5	2.9	1611	2 T38236	hypothetical prote
34	147.5	2.9	5327	2 T13564	microtubule-associ
35	146.5	2.8	670	2 T28391	ORF MSV230 hypothe
36	146.5	2.8	1254	2 JC7185	chromosome 1 Clorf
37	146.5	2.8	2910	2 T28156	DNA-directed RNA p
38	146	2.8	1271	2 A45555	glutamate rich pro
39	146	2.8	1320	2 JC5630	TCOR1 protein - mo
40	146	2.8	1435	2 S69632	regulatory protein
41	146	2.8	1791	2 T02332	hypothetical prote
42	146	2.8	3924	2 S37431	ankyrin 2, neurona
43	145.5	2.8	559	2 B44265	ENL (translocation
44	145.5	2.8	1093	2 F88556	protein B0464.5a
45	145.5	2.8	1463	2 T30290	AAS surface protei

ALIGNMENTS

RESULT 1

B84726

probable poly(ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: B84726

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; I  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.A  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84726

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-997 <STO>

A:CROSS-references: GB:AE002093; NID:g4887749; PIDN:AAD32285.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g31870

A:Map position: 2

Query Match	10.0%;	Score 517;	DB 2;	Length 997;
Best Local Similarity	32.0%;	Pred. No. 7.7e-24;		
Matches 140;	Conservative 53;	Mismatches 126;	Indels 114;	Gaps 13;
QY	569	WKVLEEAQAHLVQSILPDMVKIALCLPNI	-----CTQPIPLKQKWNHS	614
Db	79	FDELIDKESKRFDEIIPALASLLQFPESLLEVFHFQFNADNIVSGIKTGURLLNSQAGI	138	
QY	615	VTMSSQSIASLLANAPFCTFPRRNAMKSEYSSYPDINFNR--LFEGRSSRKEPKLTLF	672	
Db	139	VFLSQELIGALLACSFCLFPDDNRGAK-----HLPVNFDELHSLVLSYSQSQSKIRCM	194	
QY	673	CYFRRVTEKPTGLVTFTRQ--SLEDFPWERCEKPLTRLHVTEGTIEGNGMQLQVDF	730	
Db	195	HYFERFCSCVPIGVSEFKITAAPDADFWSKDVSL-----YQPD-----NALEVDF	242	
QY	731	ANRFVGGVGTAGLVOEIRFLINPELIVSRLFTEVLDHNECLIIITGQYSEVTVGAET	790	
Db	243	ANKYLGSGSLRGCVQOEIRFMINPELIIAGMLFLPRMDDNEATEIVGAKRFSCYITGASS	302	
QY	791	YRWARSHEDGSEKDMQRCTEIVADAL-----HFRYLDQFVPEKVRRELKAYCGF	844	
Db	303	FRPAGEYIDKKAMPFRARRTRIVADALCTPKRRHFKDIC-----LLREINKALCGF	355	
QY	845	L-----RFG-----VPSENL-854		
Db	356	LNCCKAWHQNTFMDGDNIEIQLVRNRDGLLRTTETTASHRTPLNDVEMNREKPPANNLI	415	
QY	855	-----SAVATGNWCCAFGCDARLKALIQILAAAAAERDVV-YFTFGDSE	898	
Db	416	RDFYVEGVDNEDHEDDGVATGNWCGVFGDGPDLKATIQMLAASQTRRPISYITFG-VE	474	

QY 899 LMRDIYSMHTFLTERKLD 916  
 Db 475 ALRNLDOIPTKLSSRLD 492  
 RESULT 2  
 T21138  
 hypothetical protein F20C5.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
 C:Accession: T21138  
 R:Matthews, P.  
 submitted to the EMBL Data Library, December 1995  
 A:Reference number: Z19381  
 A:Accession: T21138  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-726 <WILL>  
 A:Cross-references: EMBL:Z68161; PIDN:CAA92299.1; GSPDB:GN00022; CESP:F20C5.1  
 A:Experimental source: clone F20C5  
 C:Genetics:  
 A:Gene: CESP:F20C5.1  
 A:Map position: 4  
 A:Introns: 15/3; 162/3; 219/3; 316/3; 347/3; 403/3; 442/2; 536/3; 578/1; 675/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F20C5.1  
 Query Match 8.6%; Score 443.5; DB 2; Length 726;  
 Best Local Similarity 21.9%; Pred. No. 1.7e-19;  
 Matches 184; Conservative 124; Mismatches 276; Indels 255; Gaps 32;  
 QY 199 BEDRNDQFLTPIKLANTKPTVGDQARSNCXSGRSQVXKDCQCOQEEVDVLPESPLS 258  
 Db 14 QDEKDYEDYV-GVGFHQVPTM---KRRKLTGHGNTTESKEDPE--EPKSRDVFVSSQSS 67  
 QY 259 DVGAEIDGTGPK-----NDNKLGTQESSLGD---SPPFKES 292  
 Db 68 DESQDSAEENPEIAKEVSENCENLTETLKISNIESLDNVTSEHTLDNKHKSTPEMEDV 127  
 QY 293 EPESPMDV-----DNSNSQDSADSETSPVF-DEQD----- 324  
 Db 128 NKSNIIDVAINDEDELVLSENNKEMRDEGVQQLSQDLFADDQLLEYFGIMKDTTQ 187  
 QY 325 ---DRSSQTANKLSSCOAREADGDLKRYLTKGSEVRLHFQEGENAGTSDLNAPSG 380  
 Db 188 LDITDSEVETAKMEMIBETADS-----TFVGEDSKATKTVRTSSSS 230  
 QY 381 NSSLNVECRSKQHGKR--DSKITDHFMRISKSE-----DRKKEQCEVRQR 426  
 Db 231 FLSTVST-CEAPAKGRARMYQKELEKHVIAFTGNLTLPQDLNKKVDPDRNRYCTI--- 285  
 QY 427 TERKIPKIPNLPPEKKWLGTPPEEMRKMPGCHLPSLRPSASHTVTRVDLLRAGEV 486  
 Db 286 -----PNFPASQ---GKLREDNRVGPKVL----- 307  
 QY 487 PKPFPFTHYKLDWNKHKVMPCEQNLYPVEDENGER-----TAG--- 525  
 Db 308 ---PQRMREF-----DSRGRRRDSYFYFKRKLGLDKYKTYTGYF 344  
 QY 526 -----SWELLQIATLNKFTPQNLKDALIKYVA-----YSKKWDFTALVDVFDKVL 573  
 Db 345 MFVGLLHNWEEF-----DPDITYKLPALEMYKEMSELVGREEVLEKFA 388  
 QY 574 EEAEAOHLYQSILPDMVKIALCLPNICTQPIPLLKQKMHVSMTSQEQIATSLANAFPT 633  
 Db 389 RVARIKTAEDILPE--RIYRLVGDV-----ESATLSHKQCAALVARWFF-- 431  
 QY 634 FPRRNAMKSEYSSYPDINFNRLFEGRSRKPEKLTLCYFRVTEKPTGLTVF--TR 691  
 Db 432 -----ARPDSPS-----FCRILSSDKIGICEKLFLETFYEDKMSMDPPDGAVSFLTK 480  
 QY 692 QSLDF-PDW-ERCEKPLTRLHVYEGTIEGNGRGMLOVDPANRFVGGGVTCAGLVQREI 749

Db 481 MDKDTFNEWKDKKLSLPEVRFDEMLIEDTAL-CTQVDFANEHLGGVNLHGSVQEEI 539  
 QY 750 RFLINPELLYSLRFLTEVLIDHNECLIIITGTQYSEYTCYATYRWA-----RSHEDGSE-K 803  
 Db 540 RFLMCPBMVMYGMLLCEKMKQLEAISIVGAYVFSYSTGYGHTLKWAELEQPNHRSRQNTNEFR 599  
 QY 804 DDMQRCTEIVADALHFR-----YLDQFVPEKVRRELKAYCGELRPGVPSNLSAVA 858  
 Db 600 DRFGRLVETIADAILFKGSKLDCQTEQNLKANILIREMKKASIGFMSQGPFTNI-PIV 658  
 QY 859 TGNWGCAGFGADARLKALIOILAAAAAERDVVYFTGDSLMRDYISMHTPLTERKLDVG 918  
 Db 659 TGNWGCAGFNGDKLKFIIQVIAAGVADRLPHFCSGFGEPELAAKCKKILIERMKQKDVTLG 718  
 QY 919 K 919  
 Db 719 K 719  
 RESULT 3  
 A84726  
 probable poly (ADP-ribose) glycohydrolase [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: A84726  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayan, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: A84726  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-364 <STO>  
 A:Cross-references: GB:AB002093; NID:94887750; PIDN:AAD32286.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g31860  
 A:Map position: 2  
 Query Match 6.4%; Score 328; DB 2; Length 364;  
 Best Local Similarity 27.3%; Pred. No. 8e-13;  
 Matches 105; Conservative 54; Mismatches 123; Indels 102; Gaps 11;  
 QY 573 LEEBAQHLYSIILPDMVKIALCLPNICTQ-----PIPLLKQKMHVSMTS 618  
 Db 1 MSKESSSRWFNEFLPAMACLLLRFPSSLLESHYLSNDNLINGTKTGLRVLPNKAGIVELS 60  
 QY 619 QEQIATSLANAFCTEPRRNAMKSEYSSY-PDINFNRLF-----EGRSSRKPEKLTFL 672  
 Db 61 QELIGALLSCSFCLFP-----VDRGNSHLPINFDKLFGLSLINTGRNEHQENKIKCII 115  
 QY 673 CYFRRVTEKPTGLTVFTTRQ--SLE-----DFPEWERCEKPLTRLHVYTGTTGNGRG 724  
 Db 116 HYFQRLSSISPGFVSPERKILSLQDSSTLDEGFWGKSTVNLCPVEVRTSGLIEDQSV 175  
 QY 725 MLOYDFANRFVGGVTCAGLYOERFLINPELIVSLRFTVLDHNECLIIITGTEOYSEY 784  
 Db 176 ALEVDFAKNLGGGALRKGCVOEIRFMINFMINFVGLMFLFTMETEALFVVGARERFSLY 235  
 QY 785 TGYAETYNARSHEDGSEKDDWQRCTEIVADALHFRYLDQFVPEKVRRELKAYCGF 844  
 Db 236 TG---CFKAK-----TFIVADALR----- 253  
 QY 845 LRPGVPSNLSAVATGNWGCAGFGADARLKALIOILAAAAAERDVVYFT-----FGDS 897  
 Db 254 -HPGV-----SQYKLESLLSVLLSSGRPIRLYMGVSLQGIQGDV 293  
 QY 898 ELMRDIYSMHTFLTERKLDVGKY 921  
 Db 294 VLMVEILSSSLFFNGLRFRNSLY 317



368 HT-----TVSDITGHQS-----VTDSDE-----386  
 275 LTCQSSSLGSDPPFEKESPEPMVDNRSNQCDSSEADDETSVPFDEQDRSSQTANKL 334  
 387 ---QNNCMSTPPKIEIESIEISVVS-----KSKIESSSVSVKQNHTEKQVAIET 438  
 335 SSCQAREADGLKRYLTGSEVRLHFQFGE--NNAGTSDLNAPSGNSSSLNVCERS 392  
 439 PEQKVEKEDE-----HLNQGSEFIEESTKQPISSKPTSPDMDAATGG 484  
 393 KOHGK--RDS-----KITHFMWIKSEDRRKEQCEVRQRTKPKYIPNMLPP 441  
 485 RVSSSSPRDKILOTFSPRSTIDSFNSIKK--RNSSEANDENETNLKIP-----IP 535  
 442 EKWLGTFPIEMKMPKCGIHLPSLRSASHTVTVRVDLLRAGEVPKPPFTHYKDLMDNK 501  
 536 EKRRKFQVLSQKNI-----LVSTEDSDHEPVKVTEDSQTALHVK-----FEDL-ENK 583  
 502 HVKMPCEQNLVYV-EDENGERTAGSRWELIQTALLNKFTRPQNLKDALIKYNVAYS--- 557  
 584 SME-----SEQSLLISESNDK-----PLIDLPLLAIKRKNLNVSGVLEKKGKSTSTK 634  
 558 KKWDFLTALVDFMDKVLBEAAQHLVQSILPDMVKIALC 595  
 635 TKFD-TSIVDFIEPKTE-----ISEVLPEKRRKAIC 665

RESULT 6  
 T16420  
 hypothetical protein F52C9.8b - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
 C:Accession: T16420  
 R:Favell, T.  
 submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid F52C9.  
 A:Reference number: Z18511  
 A:Accession: T16420  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1165 <FAV>  
 A:Cross-references: EMBL:U039850; NID:g1055052; PID:g1055056; PIDN:AAA81058.1; CESP:F52C9  
 C:Genetics:  
 A:Gene: CESP.F52C9.8b  
 A:Introns: 21/1, 36/1, 61/3, 79/3, 107/3, 464/3, 566/1, 1010/3, 1029/3, 1106/3

Query Match 3.2%; Score 166; DB 2; Length 1165;  
 Best Local Similarity 19.7%; Pred. No. 0.038;  
 Matches 146; Conservative 94; Mismatches 251; Indels 250; Gaps 33;  
 32 PCGRRLVDPKAPVQFVRPPSPACV---SGRAGPHRGNAVSFVFKQKTIITWMDTKGP 88  
 478 PPQQRKQK 523  
 89 KTAESK-----ENNTRIDSMSSSVQKDNFYPHKVE 121  
 524 VAQSSSEKIEPVDKVPAPVPPQVPTPTKPTITNNKKRID-----VVTLEDAPRRVQ 579  
 122 KLENVQLN-----LDKSPTEKSSQYLNQOQTASVC-----KWO-----155  
 580 VKQEIPEVSTSDATKSDAAPTARGAVRIKQEVSDVAPNTIILSAKFERMKAEDKE 639  
 156 -----NEGHAELASPEPPAGTPLPKQ--LSNANIGQ-----186  
 640 DMKKKIALQALFNIOQERREVEKIEAFAFATTAQVQNPQPASSVQIAQVSTSESAPGT 699  
 187 -----SPTDDH-----SDTHEDRDNOQFLTPIKLANK-----PTVG 221  
 700 SEAAATMTSPKTKNNVIVETEGEQEDEDE-----IPIKSKKRAKIVNSNDEEFPVR 755  
 222 DQARSNCKS-----GSRQVKDCTGQCEQEVVLPESP-----LSDVGAEDIGTGPKN 271  
 756 HPTKRSDEKREKRVHVAESDDDMFVVKKKRRNQSPEDPEYSAASPSDEDEDDDDIGSFVS 815

272 DNKLTQBS-SLGDSPPEKESEPEPM-DVNSRNSCODSEADDETSVPFDEQDRSSQ 329  
 816 DNEDDDADSVFVGDDEPIEYEEDEDDMIERSSSRKRSRKSATPT-DRRRSRDTP 874  
 330 TANKLSSCOAREADGLKRYLTGKS-EVRLHFQFEGENNAGTSDLNAPSGNSSSLNVE 388  
 875 TQSR--SMESTSPNDRKRSRETTPGNRSMRRTSPSGRAKSRDTP--TASSSMSSSTLSY- 929  
 389 CRSSKQHGKRSKLTIDHFMWIKSEDRRKEQCEVRAHRTK- 430  
 930 CKSKETPMSEIE-----QOKKARORCKTREENRERKRLAQLELESETTGVRR 983  
 431 IPKYIPNLPPEKKLGTPIEMRKMPR-----C 459  
 984 TLRSTODNSDPLDASLATTIEEFRTKKDAKSENKAKEKQKPMNKRPTSSASVDSND 1043  
 460 GIHLPLSRPSASHTVTVRVDLLRAGEVPKPP-----PTHYKDLWNKHKVMP 506  
 1044 GVHIPAKR-----MAHASSVPGPSRSKPPMIGAVKVRPNH-TEMLDKRNKE-- 1088  
 507 CSQNLVPEDENGERTAGSRWELIQTALLNKFTRPQNL-----KDALIKYNV-----A 555  
 1089 -SEKRRKORDEL-ERLRNKKHTTEBEKI--KWARLQNALKVVGKAAGLAKATVKKELTGS 1144  
 556 YSKKWDFTALV-DFWDKVLSE 575  
 1145 PAKKQKPAFLCRKEFWISRLSE 1165

RESULT 7  
 A71623  
 probable secreted protein PF0115w - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
 C:Accession: A71623  
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
 Science 282, 1126-1132, 1998  
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; MUID:99021743; PMID:9804551  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1192 <GAR>  
 A:Cross-references: GB:A8001373; GB:AE001362; NID:g3845097; PIDN:AACT1813.1; PID:g38450;  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PF0115w

Query Match 3.2%; Score 165; DB 2; Length 1192;  
 Best Local Similarity 21.5%; Pred. No. 0.045;  
 Matches 85; Conservative 61; Mismatches 143; Indels 106; Gaps 18;  
 70 TSFVFKQKTIITWMD-TKGPKTAESEKNNTRIDSMSSSVQKDNF-YPHKVEKLEN-V 126  
 235 SDFNMKMYQNTQIDNNKGGKTTTMEKNKKNRDKKSKKRTKONKYKKENQNIENHI 294  
 127 POLN-----LDKSPTE-KSSQYLNQOQ-----TASVC-----KWNQEGKHAQLL 165  
 295 PQSKYQERIEILDNGKELSKHKIKKEGIEKTDITNIADIKKEERTKDEKKN 354  
 166 ASEPPAGTPLPQLSNANIGQSPH-----TDDHSDTHEDRDNOQFLTPIKLANT 216  
 355 IQQ-----LVKDVQLIKVGEETKDEKEDKEDTDEEDTDEEDTDEEDTDEEDTSD 408  
 217 KPTVGGQARSNCKSGSRQSKVDCGQCEQEVVLPESPPLSDVGAEDIGTGPKNDKLT 276  
 409 EETIGDQE-----NKEETEVE-----DEKKT 428  
 277 --QESLSDGSPPEKESPEPMVDNRSNQCDSSEADDETSVPFDEQDRSSQTANKL 334  
 429 EKAEELLEDKESEKDK-ESEKDESEKDESEKDEKKT-----EEDENKEDT-EKG 482

QY 439 DNKHKVPCBONLVPVDENGERTAGSRWELIQTALLNKFTTRPQNLKDALILKINVAYSK 558  
 Db 698 PEHFALVPLTESQ-GPHSGSGSRTSGCRQAVVVOEDSRKORLPLPLRDT----- 746  
 QY 559 KWDFTALVDVFDWQVLEAEQAHLQYSLIPDMVKIALCLPNICTQP----- 603  
 Db 747 -----KLLSPLRDTPPQSL-----MVKITLDDLRLSRIPQPPGKSRQKAEKQ 791  
 QY 604 -----IPILKQRM-----NHSVTM-----SQEIASILLANAFCTF 634  
 Db 792 PAGKHSSEKSSSSSLAKKRGAEARDCNKKIRLEKEIKSSQSSSSSHKSSKTK 851  
 QY 635 PRNNAKSEYSSYPDINFNRLFEGESSKPEK-LKTLFCYFRRVTE---KKPTGLVTF 689  
 Db 852 PSRPSQSSKEMLPFPVS-----SSQKPAKALKR-----SRRADTCGDPFKSASS 902  
 QY 690 TQOSLED--FPWERCEKPLRLHVTYEGTI-----EGNGR-GMLQVDFANR 733  
 Db 903 TSNHDKSSIPKORRVEGKSSSSSEKSSGSGDTANFPVPSPNGSKGKQVAFDKQ 962  
 QY 734 FVGGVGTGAGL-VOEBIRFLINPELIVSR-----FTEVLHNECLIIITGEYSE--- 783  
 Db 963 -----QADLHMRKAKMKQKAEIMTDRVGKAFKYLEAVLSFIEGIIATESQSSKSA 1015  
 QY 784 YTGAYET-----YRWARSHEDGSEKDDWQ-----RRCTEIVADIALHRRVLDQFVPE 831  
 Db 1016 YSVSETVDLIKFMISLFSFSDATPQEKIFAVLCMRCOSI--LNMWFRCKD--IAI 1071  
 QY 832 KVERELNKACGFLRCPVPSNLSAVATGNWCGAFGG 869  
 Db 1072 KYSRTLNKHF-----ESSKVAQAPSPCIATSG 1099

RESULT 9

T06029  
 hypothetical protein T28119.100 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06029  
 R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: 215484  
 A:Accession: T06029  
 A:Molecule type: DNA  
 A:Residues: 1-532 <BEV>  
 A:Cross-references: EMBL:AL035709; GSPDB:GN00662; ATSP:T28119.100  
 A:Experimental source: cultivar Columbia; BAC clone T28119  
 C:Genetics:  
 A:Gene: ATSP:T28119.100  
 A:Map position: 4

Query Match 3.2%; Score 163; DB 2; Length 532;  
 Best Local Similarity 23.1%; Pred. No. 0.019;  
 Matches 83; Conservative 61; Mismatches 145; Indels 70; Gaps 17;  
 QY 92 ESEKNNNTRIDSMSSVQKDNFPHKVEKLVNLPQLNLDKSPTEKSSQYLNQOQTASV 151  
 Db 189 EGESKSNSTENV-----SVHEDESGP-KNEVLEGSVIVEKSVLNTTENGSDGGEQETKS- 241  
 QY 152 CKWQEGKHAQQLLASEPPAGTFLPK-QLSNANI-----GOSPHITDDHSDTDH 198  
 Db 242 ---ELDSKTGEGFSD---SNGELPETNLSTSNATETTESGSDSGSGSKSTGYQTKN 295  
 QY 199 BEDRDNQOFLTPIKLANTKFTVGDQCARSNCKSGRSQVXDCTCCQEEVDVLPESPLS 258  
 Db 236 EDE-----KEKVSSEESKVKESGKNE--KDASSQDESEKPERKKK 339  
 QY 259 DVGAEIDGTGPKNDNKLTCQESSLGDSPFPFEKESPEPMVDNRSNCSQDEADEETSP 318  
 Db 340 EESSQ-GEKKEEPKPEKEDSSS-----QESKEEPEPENKEKASSQEEINEIKETE- 392  
 QY 319 VFDEQDSSSTANTKLSSQCAEADGLRKRYLTGKSEVRLHFQFEGENAGT----- 371

QY 335 SSCQAREADGLRKRYLTGKSEVRLHFQFEGENAGTSDLNAPKSGNSSLNVECESSKQ 394  
 Db 483 TEVIKKETDVEKKE-----KGYEGTDEEDKEKEEDDE-----ETKVERK 525  
 QY 395 HGKRDSTKTDHFWRIKSDRRKKEOCEVRHQRTER 429  
 Db 526 KTEKDEGTDEEDTDDSD--KDEETKVEEKTER 558  
 RESULT 8  
 I39410  
 AF-4 protein, splice form 3 - human  
 N:Alternate names: serine/proline-rich FEL protein  
 C:Species: Homo sapiens (man)  
 C>Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 29-Aug-1997  
 C:Accession: I39410  
 R:Nakamura, T.; Alder, H.; Gu, Y.; Prasad, R.; Canaanani, O.; Kamada, N.; Gale, R.P.; Lang  
 Proc. Natl. Acad. Sci. U.S.A. 90, 4631-4635, 1993  
 A:Title: Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leuk  
 A:Reference number: A47440; MUID:93281633; PMID:8506309  
 A:Accession: I39410  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <NAK>  
 A:Cross-references: GB:I13773; NID:g306446; PID:g306447  
 C:Comment: This protein is one of several to form chimeric fusion proteins with the tri  
 C:Genetics:  
 A:Gene: GDB:MLLT2; AF-4; FEL  
 A:Cross-references: GDB:I136792; OMIM:159557  
 A:Map position: 4q21-4q21  
 C:Keywords: alternative splicing

Query Match 3.2%; Score 163.5; DB 2; Length 1210;  
 Best Local Similarity 19.7%; Pred. No. 0.057;  
 Matches 197; Conservative 112; Mismatches 368; Indels 321; Gaps 42;  
 QY 35 QRRVLDPKDAPVQFRVPPSSPACVSGRAGPHRGATSFVFKQKTIITWMDTKGPKTAESE 94  
 Db 260 QDSIVAQAPPSQTFPPPSLP-----SKVAMQOKPTAYVRMDGQDQAPSE 306  
 QY 95 SKENNTRIDSMSSVQKDNFPHKVEKLVNLPQLNLDKSPTEKSSQYLNQOQTASVCKW 154  
 Db 307 SPELKLPEYRQOTFEKIDL---KVPKAKILTKMPSQSEVQT-----Y 349  
 QY 155 QNEGKHAQLL-----ASPPAGTFLP----- 176  
 Db 350 SNEVHCVEILKEMTHSWPPPLTALHTPSTAEPKFPFPTKDSQHVSVTVQNKQYDTSS 409  
 QY 177 KOLGNANTGQSPHTDDSDTHEDRDNQOFLTPIK--LANTKPTV-----G 221  
 Db 410 KTHNSQOQTSMLEDDLQLSDSEDSDEQ--TPEKPPSSAPPAPQSLPEPVASAHSS 467  
 QY 222 DQARSNCKSGRSQVXDCTCCQEEVDVL-----PESPLSDVGAEDIDGTGPKNDKL 275  
 Db 468 SAESSESTSDSSSSSESSSSSESENEPLETPAPEPEPTTNK-----QLDNWL 519  
 QY 276 TQESSLGDSPFPFEKESPEPMVDNS-----RNSCQSEADETSPVFDQDRRSQT 330  
 Db 520 T-----KVSQPAAPPEGRSTEPFRPHRPESKSSDSATSOHSESKDPPPKS 566  
 QY 331 ANKL-----SSQCAEADGLRKRYLTGKSEVRLHFQFEGENAGTSDLNAPK 378  
 Db 567 SSKAPRAPPEAPHGKRSQKSPAQEPQPR-----QTVGKQ-PPKP 608  
 QY 379 SGNSSLNVECSKQHGKRDSDTHFWRIKSDRRKKEOCEVRHQRTERIKPIYPN 438  
 Db 609 V--KASARAGSTSLQ-GEREPGLLYGSRDQTSKDKPKVKYKGRPRAASNEPK---PA 662  
 QY 439 LPPEKKWLGTPIBEMKRPKRCGTHLPSLRPSASHVTTVVVDLLRAGEVPKPTTHYKDLW 498  
 Db 663 VPPSE-----KKHKSSLPAPS-----KALSGEPKADNVEDRT 697

1800 SEP-----SPMEEFVKLETSPV-PKEEPIKMEESPEQTPTPDLLISNNESQDTPGAVNNHL 1853  
495 KOLMD-----NKH-VYMPGSEONLYPVDENGERTAGSRWELIQTALLNKTRP 542  
1854 HENHDAVQTPIQLOPASQHQVAQSPRAVADPSQONG-----PVLVSQSQSP 1901  
543 QML--KDAILKYNVAYSKKNDFTALVDVDFWVKVLEAEA-----QHLV-----582  
1902 SPMSQSQSDMAQNLTLSK-----DINDLAALKHKHPEALQAQTRGDCSGIFQHLLHAQG 1957  
583 ---QSILPDWVKIALCLPNICTOPIELLKQKMNHVTMSQEQIASILLANAFCTFPERNAK 640  
1958 NGQWTPMELQKAAF-----FAQQQENANQMWO-----AK 1989  
641 MKSEYSSYPDINPNRUFEGRSRKEKLTLCYFRRVTEKKPTGLVTFTROSLEDFPEW 700  
1990 MKQQ-----TINKDRI-----KEQFRVMRYEENRKEVE-----EDRREK 2024  
701 ERCEKPLTRL 710  
2025 QRKEERQRL 2034

RESULT 11  
E88320  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: E88320  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99049613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an  
A:Accession: E88320  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2738 <STO>  
A:Cross-references: GB:chr\_II; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A1  
C:Genetics:  
A:Gene: F07A11.6  
A:Map position: 2

Query Match 3.1%; Score 160; DB 2; Length 2738;  
Best Local Similarity 19.4%; Pred. No. 0.3;  
Matches 132; Conservative 104; Mismatches 235; Indels 210; Gaps 31;

QY 84 DTGKPKTAESKKNNTRIDSMMSVQKDFYHKVEKLENVQNLNLDKSPTEK-----138  
DB 1526 DSSSGSSDSSSSDGSSTSSDSS-----SDDEVPAKQTEPVTISIPVASDNGSPENVVET 1579  
QY 139 ---SSQYLNQOQTASVCKWQNEKGKAEOLLASEPAGTFLPKQLSNANIG--QSPTDDH 193  
DB 1580 PSIVSQTPREPFPFTISEQSS--SEPEAVPECEASVEFQMETISQNVPEVSEHEDSH 1636  
QY 194 STDHEDHDNQOFLTPIKLANTKTVDGGQARSCKSGSRQSVKDKTCGQOQEEVDLP 253  
DB 1637 EHGDSVAVESQO--QPLE-----HQEKEEL- 1661  
QY 254 EPLSDVGAEDIGTGPKNDKNKLTGQESSIGDS-PPFEKESPESPMDVNSRNSQDSEA 312  
DB 1662 ENKILDVAE-----HHEEQVQGEDSVESSEIPAPSDPEPVTQAEKSAHTLISDQET 1715  
QY 313 DEETSPVFQDQ--DRSSQTANKLSSQAEAGDLKRYLTGKSEVRLHFQEGENNAQT 371  
DB 1716 DQAVQISFDEEAEDEFPQYDPFGISTNEKESVG--KOPHNIRPTE-----PLNNGH 1764  
QY 372 SDLNAPSGNSSLNVCSSKQHGKRDKSITDHFMRISKESDRRKEQCEVHRQTERKI 431  
DB 1765 TDLIFGSSSAHA-----SEKGS-----TKSEDDMEDSELVMEKEVPM 1804



177	QY	KOLSNANI	QOSPH	DDHSD	TDHEED	RNQOFL	TFIK--	LANTK	TPV-----	G	221																																						
417	DB	KTHNS	QOGT	SMLEDD	LQ	SDSE	SDSEQ--	TPKPP	SSAPPASQ	SLPEP	VASAHSS	474																																					
222	QY	DGOAR	SNCK	SGSR	QSV	KDCT	CGOQ	EEVDVL-----	PESPL	SDVCAE	DIGT	CPKDNKL	275																																				
475	DB	SAESES	TDSD	SSDSE	SSSE	SSDSE	SEEN	PLET	PAPEPE	PPPTNKW-----	QDNWL	526																																					
276	QY	TGOESS	SLGDS	PFPE	KESPE	SPMD	VDNS-----	RNSCQ	SEADE	ETSPVF	DEQDRSS	QOT	330																																				
527	DB	T-----	KVSQ	PAAP	PEGR	STEP	PRRH	PE	SGSD	SATS	QEHSE	KDP	PPKS	573																																			
331	QY	ANKL-----	SSCQ	AREAD	LKR	YLT	KGSE	VRLLHF	QFEG	ENNAG	HS	DLNAKP	378																																				
574	DB	SSKAP	RAPE	APH	PKCR	SCQ	SPAQ	QEPQ	R-----	QTVG	TQ--	PKPK	615																																				
379	QY	SGNSS	LN	VECR	SSKOH	GKDS	KITD	HFMR	ISKSD	RRKEQ	CEVR	HTER	KLIP	KYIP	PN	438																																	
616	DB	V--	KASAR	AGRT	SLQ--	GGRE	GLLP	YG	SRD	TSKDK	PKV	TKGR	PRAA	SNFK--	PA	669																																	
439	QY	LPBKK	WLG	TPIE	MRK	PRC	GIH	PLSP	LRPS	ASH	TVT	VRDL	LRAGE	VPK	PPHY	KDLW	498																																
670	DB	VPPSE-----	KXK	KS	SL	PAPS-----	KALSG	PE	PAK	DN	VEDRT	704																																					
499	QY	DNKH	VMP	PCSE	QNL	YVE	DENG	ERTAG	SRWEL	IQTALL	NKFT	RPQNL	KOAIL	KYN	VAYSK	558																																	
705	DB	PEFAL	VPL	TESQ--	GPHSG	SGRT	SGCQ	QAVV	QEDSR	KDR	LPL	PLUR	T-----	753																																			
559	QY	KWDFT	AL	VDFW	DKVL	EAB	QAHL	YQIL	PD	MV	KIAL	CL	PNIC	TQP-----	603																																		
754	DB	-----	KLLS	PL	RD	TPPQ	SL--	MVKI	TL	DL	LSRI	PQP	CKGSR	KAR	KQP	798																																	
604	QY	-----	IPLL	KQK	-----	NH	VTM-----	SOEQ	IAS	LLA	NA	FCTP	634																																				
799	DB	PAGK	HS	SEK	SRD	SSSK	LAKR	KGAE	RD	CN	KKIR	LEKE	IKS	QSSS	SSSH	KESSTK	858																																
635	QY	PRNA	KM	SE	YSY	DDI	NF	NRL	PE	GR	SR	KPEK--	LKTL	FCY	FR	RVTE--	KKPT	GLVTF	689																														
859	DB	PSR	SSQ	SSK	EM	LPPP	PVS-----	SSSQ	KAP	AL	KR-----	SREAD	T	CGD	PP	KVP	AV	909																															
690	QY	TRQSL	ED--	FP	WER	CEK	PL	RLH	VY	EGTI-----	EGNR	GML	QV	D	PANR	733																																	
910	DB	PRVNH	KD	SSI	PQ	RR	VEG	KSR	SSAD	KG	SGD	TAN	PPV	PSL	PNG	NSK	PKPQ	VK	FDKQ	969																													
734	QY	FVG	GT	GAGL--	VOE	IR	F	LIN	PEL	IV	SRL-----	PTEVL	DH	NE	CLIT	TEQ	YSE---	783																															
970	DB	-----	QADL	EM	RE	AK	MM	KQ	KAL	MT	DR	VG	KAF	KYLE	VL	SF	NEC	G	I	ATES	SSQ	SKSA	1022																										
784	QY	YTG	YAE	T-----	YR	W	A	R	G	H	E	D	G	S	E	K	D	D	WQ-----	RRCTE	I	V	A	I	D	A	L	H	F	R	I	L	D	Q	F	Y	P	E	831										
1023	DB	YSV	Y	SET	V	D	L	I	K	F	I	M	S	L	K	F	S	D	A	T	P	Q	E	I	F	A	V	L	C	M	R	C	Q	S	I--	L	N	W	A	M	F	R	C	K	D--	I	A	I	1078
832	QY	KVR	EL	N	K	A	C	G	F	L	R	E	G	V	P	S	E	N	I	S	A	V	A	T	G	N	W	C	864																				
1079	DB	KYS	R	T	L	N	K	H	F-----	ESSK	K	A	Q	A	P	S	P	C	1101																														

RESULT 14  
A47328  
natural killer cell tumor-recognition protein - human  
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumor  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
C:Accession: A47328  
R:Anderson, S.K.; Guller, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.  
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993  
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.  
A:Reference number: A47328. MUID: 93113824; PMID: 8421688

A;Experimental source: NK killer cells from adult blood  
A;Note: sequence extracted from NCBI backbone (NCBIN:122798, NCBIIP:122800)  
C;Genetics:  
A;Gene: NKTR  
A;Cross-references: GDB:137171; OMIM:161565  
A;Map position: 3p23-3p21  
C;Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology  
C;Keywords: alternative splicing; lymphocyte  
F:60-230/Domain: cyclophilin homology <CYP>

RESULT 15

RESOL.  
KF805

KFB05  
coagulation factor V precursor - bovine

coagulation factor V precursor - bovine

C/Species: Bos primigenius taurus (cattle)

C;Date: 04-Mar-1993 #sequen





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:20:49 ; Search time 15.5091 Seconds  
(without alignments)  
3249.964 Million cell updates/sec

Title: US-09-302-812-6  
Perfect score: 5164  
Sequence: 1 MSAGPGEPTKRWGAAGT.....YHVESSAETDMPGQKAGT 968

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	176	3.4	6632	1 UN89 CAEEL	O01761 caenorhabdi
2	171.5	3.3	1647	1 P0B1 CAEEL	Q10124 caenorhabdi
3	170	3.3	997	1 B1R1 SCHPO	O14064 schizosacch
4	163.5	3.2	1210	1 AFA HUMAN	P51825 homo sapien
5	162	3.1	5038	1 PCLO MOUSE	O9YX77 mus musculu
6	161	3.1	5147	1 PCLO HUMAN	Q9Y6V0 homo sapien
7	159	3.1	687	1 D5PP RAT	Q62598 rattus norv
8	159	3.1	1462	1 NKCR HUMAN	P30414 homo sapien
9	157.5	3.0	2211	1 FA5 BOVIN	Q28107 bos taurus
10	156	3.0	1274	1 ENAM MOUSE	O55196 mus musculu
11	154	3.0	934	1 D5PP MOUSE	P97399 mus musculu
12	153	3.0	5085	1 PCLO RAT	Q9JX86 rattus norv
13	152.5	3.0	406	1 SR40 YEAST	P32583 saccharomyc
14	152.5	3.0	5120	1 PCLO CHICK	Q9PU36 gallus gall
15	152	2.9	1359	1 ATRX CAEEL	Q9U7E0 caenorhabdi
16	150	2.9	489	1 DMP1 RAT	P98193 rattus norv
17	150	2.9	699	1 NP14 HUMAN	Q14578 homo sapien
18	150	2.9	1070	1 PVDR PLAVS	P22290 plasmodium
19	150	2.9	1337	1 DEXT STRDO	P39653 streptococc
20	149.5	2.9	513	1 DMP1 HUMAN	Q13316 homo sapien
21	149.5	2.9	2314	1 AKG6 RAT	Q9WVC7 rattus norv
22	149.5	2.9	1453	1 NKCF MOUSE	P30415 mus musculu
23	147.5	2.9	1358	1 SIRA YEAST	P11978 saccharomyc
24	147.5	2.9	5596	1 MDN1 HUMAN	Q9NM22 homo sapien
25	146	2.8	1395	1 SP41 YEAST	P38904 saccharomyc
26	146	2.8	3924	1 ANK2 HUMAN	Q01484 homo sapien
27	145.5	2.8	559	1 ENL HUMAN	Q03111 homo sapien
28	145.5	2.8	1093	1 YKD5 CAEEL	Q03563 caenorhabdi
29	145.5	2.8	1282	1 BMS1 HUMAN	Q14592 homo sapien
30	145.5	2.8	1447	1 BUD4 YEAST	P47136 saccharomyc
31	145	2.8	574	1 MIG1 CANAL	Q9Y7G2 candida alb
32	143.5	2.8	503	1 DMP1 MOUSE	O55188 mus musculu
33	143	2.8	1253	1 D5PP_HUMAN	Q9NZW4 homo sapien

RESULT 1  
UN89 CAEEL  
ID UN89 CAEEL STANDARD; PRT; 6632 AA.  
AC O01761; Q17362;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
GN UNC-89 OR C09D1.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
RC STRAIN=Bristol N2;  
RX MEDLINE=96180278; PubMed=8603916;  
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;  
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
assembly, encodes a giant modular protein composed of Ig and signal  
transduction domains.";  
RT J. Cell Biol. 132:835-848 (1996).  
RL [2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Du Z., Le T.T., Wilson R.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP REVISIONS.  
RA Waterston R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RL Submitted (APR-2002) to the muscle M-line. Myofibrillar  
lattice assembly begins with positional cues laid down in the  
basement membrane and muscle cell membrane. UNC-89 responds to  
these signals, localizes, and then participates in assembling an  
M-line  
CC TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
CC -! SIMILARITY: Contains 1 DBL-homology (DH) domain.  
CC -! SIMILARITY: Contains 1 fibronectin type III domain.  
CC -! SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
CC -! SIMILARITY: Contains 1 PH domain.  
CC -! SIMILARITY: Contains 5 RCSD domains.  
CC -! SIMILARITY: Contains 1 SH3 domain.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U33058; AAB00542.1; --  
CC EMBL; AF003131; AAB54132.2; --  
CC FDB; 1FHO; 20-DEC-00.  
CC WormPep; C09D1.1; CE30426.

34 141.5 2.7 1489 1 YGP0 YEAST  
35 141 2.7 1031 1 YDG9 SCHPO  
36 140.5 2.7 646 1 SGI BOVIN  
37 140.5 2.7 1087 1 POMI SCHPO  
38 140.5 2.7 3259 1 GOBI HUMAN  
39 140 2.7 623 1 HPC2 YEAST  
40 139.5 2.7 1088 1 MRTE HUMAN  
41 139.5 2.7 1176 1 KMLS BOVIN  
42 139.5 2.7 2738 1 PGLV RAT  
43 139.5 2.7 5560 1 SPEN DROME  
44 139 2.7 925 1 FIPI YEAST  
45 139 2.7 950 1 STL2 MOUSE

P53115 saccharomyc  
Q10496 schizosacch  
P23389 bos taurus  
Q09690 schizosacch  
Q14789 homo sapien  
Q01448 saccharomyc  
Q9ULH7 homo sapien  
Q28824 bos taurus  
Q9ERB4 rattus norv  
Q8AX83 drosophila  
P40020 saccharomyc  
Q99N50 mus musculu

ALIGNMENTS





[illegible]

15-JUL-1998 (Rel. 36, last sequence update)  
10-OCT-2003 (Rel. 42, last annotation update)  
DE Bir1 protein (Chromosome segregation protein cut17).  
GN BIR1 OR CUT17 OR PBH1 OR SPC92.02C OR SPC31B10.10C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
ON NCBI\_TaxID=4896;  
RX [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
RX MEDLINE=21439264; PubMed=11554922;  
RA Morishita J., Matsuoka T., Goshima G., Nakamura T., Tatebe H.,  
RA Yabuta M.,  
RA "Bir1/Cut17 moving from chromosome to spindle upon the loss of  
RA cohesion is required for condensation, spindle elongation and  
RA repair.";  
RL Genes Cells 6:743-763 (2001).  
RX [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaigo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welljens I., Vanstreels E., Risger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Kogej P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gailland C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.R., Moreno S., Armstrong J., Forsberg S.L.,  
RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potaashkin J.,  
RA Shekovskii G.V., Uesery D., Barrall B.G., Nurse P.;  
RL "The genome sequence of Schizosaccharomycetes pombe.";  
RL Nature 415:871-880 (2002).  
RX [3]  
RP CHARACTERIZATION.  
RX MEDLINE=99398681; PubMed=10458581;  
RA Uren A.G., Beilharz T., O'Connell M.J., Bugg S.J., van Driel R.,  
RA Vaux D.L., Lithgow T.;  
RA "Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell  
RA division.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:10170-10175 (1999).  
RX [4]  
RP CHARACTERIZATION.  
RX MEDLINE=21850422; PubMed=11861551;  
RA Rajagopalan S., Balasubramanian M.K.;  
RA Schizosaccharomycetes pombe Birp, a nuclear protein that localizes to  
RT kinetochores and the spindle midzone, is essential for chromosome  
RT condensation and spindle elongation during mitosis.";  
RL Genetics 160:445-456 (2002).  
RX [5]  
RP FUNCTION.  
RX MEDLINE=20035862; PubMed=10571085;  
RA Rajagopalan S., Balasubramanian M.K.;  
RA "S. pombe Phbp, an inhibitor of apoptosis domain containing protein  
RT is essential for chromosome segregation.";  
RL FEBS Lett. 460:187-190 (1999).  
CC -/- DIVISION. Has a role in chromosome segregation by recruiting

condensin and ark1 kinase to appropriate sites as the cell progresses through mitosis.

-!- SUBCELLULAR LOCATION: Nuclear. Interacts with the outer centromeric regions of the chromosomes during interphase. After chromatid separation moves to the middle of the spindle.

-!- SIMILARITY: Contains 2 BIR repeats.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AB031034; BAA83415.1; -  
 EMBL; AL031323; CAA20434.1; -  
 EMBL; AL121859; CAB58376.1; -  
 PIR; T43523; T43523.  
 HSP; Q13490; IQBH.  
 GeneDB SPombe; SPC962.02c; -  
 InterPro; IPR001370; BIR.  
 Pfam; PF00653; BIR; 2.  
 SMART; SM00238; BIR; 2.  
 PROSITE; PS01282; BIR REPEAT 1; FALSE\_NEG.  
 PROSITE; PS0143; BIR REPEAT 2; 2.  
 Cell division; Mitosis; Nuclear protein; Repeat.  
 REPEAT 25 99 BIR 1.  
 REPEAT 120 194 BIR 2.  
 DOMAIN 80 83 POLY-ASP.  
 DOMAIN 312 319 POLY-ASP.  
 DOMAIN 487 490 POLY-SER.  
 SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489F4 CRC64;

Query Match 3.3%; Score 170; DB 1; Length 997;  
 Best Local Similarity 21.2%; Pred. No. 0.021;  
 Matches 135; Conservative 82; Mismatches 227; Indels 194; Gaps 32;

7 WEP-----CTKARGMAGTSAPTASDSRSPFGRRVLDPKDAPVQFRVPP 52  
 173 WEPPDDPYTEKRRADCVFTWKDPNSLSPT---KLSFLSTN-IDPEDLTEDNSILP 227  
 53 SSPACVSGRAGPHRGNATSVFKQKTTTWMYTKGPKTAE-SESKENN-NTRIDSMW--- 107  
 228 VSP-----TRDSTKSHKTLNFSRKNLNARPLTWSLYT 262  
 108 -SSVQKNFPHVKLENVLPOL-----NLDKSPTEKSSQVLNQOQASVCK-----WQN 156  
 263 NTSEKDS-QPTAPQSPFKPVLPTAPRRKNKSP-----KSKPAVFKPKVPFSD 312  
 157 EGKHAECOLLASEPPAGTLPKOLSN--ANIGQSPHTDDSDTHDEDRDNOQFLTPKLA 214  
 313 EDEDDDLTASQ-----PFGKIGCNDMSQVAKNFTEIPLKEDKDEKNEHLVSPATSV 367  
 215 NTKPTVGDGQARSNCCKSGRSQSVKCTGCGQEEVDVLPESPLSDVGAEDGTGPKNDK 274  
 368 HT-----TVSDITGHQS-----VTDESDE----- 386  
 275 LTQESSLGDSPPPEKSESPMDVNSRNSCQDSEADETSPVFOEDDRSSQTANKL 334  
 387 ---QNNCMSTPPEKIEESISVSVS-----KSKEISSSVSSVSGKEQNHTEKQVALET 438  
 335 SSCQAREADGLRKRYLTKGVRLHFQFEGE--NNAGTSDLNAPKPGNSSSLNVECKSS 392  
 439 PEQKVEKDE-----HLNLQGSFIEESTKQPISSKPSSTSPDMDAATGG 484  
 393 KQHGK--RDS-----KITDHPMRIKSEDRRKEQCEVHQHTEKIPKYPNPLP 441  
 485 RVSSSSFRDKILQTFNSPRSTIDSFSNISK--RNSEANDENDETNLKIP-----IP 535  
 442 EKKWLGTPIEEMKMRPGCIHLPSLRPSASHVTYVRLLRAGEVPKPPHYKDLWDNK 501  
 536 EKKRKFQEVLSQKNI-----LVSSTEDSHEPVKVTDSQTAHVSK-----FEDL-ENK 583

QY 502 HVKMPCEQONLVPV-EDENGERTAGSRWELIQTALLNKFTRPQNLKDAILKYNVAYS--- 557  
 DB 584 SME---SEQSLISESNDK-----PLDILPLLAIKKDLNVLGVLEKGTSTSK 634  
 QY 558 KKWDFTALVDFWQKVLBEAEQAHLVQSILPDMVKIALC 595  
 DB 635 TKFD-TSIVDFTEPKTE-----ISEVLPEKKAIC 665

RESULT 4  
 ID AF4 HUMAN STANDARD; PRT; 1210 AA.  
 AC P51825;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE AF-4 protein (Proto-oncogene AF4) (FEL protein).  
 GN MLLT2 OR AF4 OR FEL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93281633; PubMed=8506309;  
 RA Nakamura T., Alder H., Gu Y., Prasad R., Canaani O., Kamada N.,  
 RA Gale R.P., Lange B., Crist W.M., Nowell P.C., Croce C.M.,  
 RA Canaani E.;  
 RT "Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in  
 RT acute leukemia share sequence homology and/or common motifs";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:4631-4635(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93184301; PubMed=8443374;  
 RA Morrissey J., Tkachuk D.C., Milatovich A., Francke U., Link M.,  
 RA Cleary M.L.;  
 RT "A serine/proline-rich protein is fused to HRX in t(4;11) acute  
 RT leukemias";  
 RL Blood 81:1124-1131(1993).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -!- DISEASE: Involved in acute leukemias through a chromosomal  
 CC translocation t(4;11)(q21;q23) that involves MLLT2 and MLL/HRX.  
 CC The result is a rogue activator protein.  
 CC -!- SIMILARITY: Belongs to the AP4 family.  
 CC -!- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;  
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/AF4.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L13773; AAA58360.1; -  
 EMBL; L25050; AAA36642.1; -  
 PIR; A58198; A58198.  
 PIR; I39410; I39410.  
 Gene; HGNC:7135; MLLT2.  
 MIM; 159557; -  
 GO; GO:0003700; F:transcription factor activity; TAS.  
 InterPro; IPR007797; AF-4.  
 Pfam; PF05110; AF-4; 1.  
 Nuclear protein; Chromosomal translocation; Proto-oncogene.  
 DOMAIN 483 492 POLY-SER.  
 DOMAIN 835 843 POLY-SER.  
 DOMAIN 866 869 POLY-PRO.  
 DOMAIN 871 874 POLY-SER.  
 DOMAIN 46 46 K -> R (IN REF. 2).  
 CONFLICT 46 46 E -> G (IN REF. 2).  
 CONFLICT 624 624 SASSTKS -> VPAVPRV (IN REF. 2).  
 CONFLICT 899 905

Db 1016 YSVISYETDLIKFIMSLKSFSDATAPTQEKIFAVLCMRCQSI--LNMAFMFRCKD--IAI 1071  
 QY 832 KYRRELNKAYCGFLRPGVPSENSAVATGNMGCGAFGG 869  
 Db 1072 KYSRTLNKHFF-----ESSSKVAQAPSPCIASTG 1099

RESULT 5  
 PCLO\_MOUSE  
 ID PCLO\_MOUSE STANDARD; PRT; 5038 AA.  
 AC Q9QYX7; Q9QYX6; Q9QZT0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Piccolo protein (presynaptic cytomatrix protein) (Aczonin) (Brain-  
 DE derived HLMN protein).  
 GN PCLO OR ACZ.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING,  
 RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.  
 RC TISSUE=Brain;  
 RX MEDLINE=99439764; PubMed=10508662;  
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
 RA Kilmann M.W.;  
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
 RT zones, shares homology regions with rim and bassoon and binds  
 RT profilin.";  
 RL J. Cell Biol. 147:151-162 (1999).  
 RN [2]  
 RP REVISIONS.  
 RA Kilmann M.W.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 4502-4692 FROM N.A.  
 RC TISSUE=Brain;  
 RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP INTERACTION WITH RIMS2.  
 RX MEDLINE=22384373; PubMed=12401793;  
 RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,  
 RA Sasaki T., Tajima N., Iwanaga T., Seino S.;  
 RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of  
 RT CAMP-GREII.Rim2.Piccolo complex in CAMP-dependent exocytosis.";  
 RL J. Biol. Chem. 277:50497-50502 (2002).  
 CC -!- FUNCTION: May act as a scaffolding protein involved in the  
 CC organization of synaptic active zones and in synaptic vesicle  
 CC trafficking.  
 CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.  
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
 CC synaptic junctions.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9QYX7-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9QYX7-2; Sequence=VSP 003928, VSP 003929;  
 CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in  
 CC stomach, testis and pancreas.  
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
 CC phospholipids. Calcium binds with low affinity but with high  
 CC specificity and induces a large conformational change.  
 CC -!- SIMILARITY: Contains 2 C2 domains.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation





RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,  
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield A.S., Krywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 [5]  
 RP SEQUENCE OF 4405-5147 FROM N.A.  
 RA Kaizicki J., Elliott G.; to the EMBL/GenBank/DBJ databases.  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May act as a scaffolding protein involved in the  
 CC organization of synaptic active zones and in synaptic vesicle  
 CC trafficking (By similarity).  
 CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
 CC synaptic junctions (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=Q9Y6V0-1; Sequences=Displayed;  
 CC Name=2;  
 CC IsoId=Q9Y6V0-2; Sequences=VSP\_003923, VSP\_003924, VSP\_003925,  
 CC VSP\_003926, VSP\_003927;  
 CC Note=No experimental confirmation available;  
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
 CC phospholipids. Calcium binds with low affinity but with high  
 CC specificity and induces a large conformational change.  
 CC -!- SIMILARITY: Contains 2 C2 domains.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; Y19188; CAB60727.1; -;  
 DR EMBL; AC004903; AAD20936.1; -;  
 DR EMBL; AC004886; AAD21789.1; -;  
 DR EMBL; AB011131; BAA25485.1; -;  
 DR EMBL; BC001304; AAH01304.1; -;  
 DR EMBL; AC004082; BAB97937.1; -;  
 DR FIR; T00634; T00634.  
 DR HSSP; P04410; 1A25.  
 DR Genew; HGNC:13406; PCLO.  
 DR MIM; 604918; -;  
 DR GO; GO:0005856; C:cytoskeleton; NAS.  
 DR GO; GO:0045202; C:synaptic junction; ISS.  
 DR GO; GO:0005509; F:calcium ion binding; ISS.  
 DR GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.  
 DR GO; GO:0005522; F:profilin binding; ISS.  
 DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
 DR GO; GO:0016080; P:synaptic vesicle targeting; ISS.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR001565; Synaptotagmin.  
 DR PRINTS; PR00360; C2DOMAIN.  
 DR PRINTS; PR00399; SYNAPTOTAGMIN.  
 DR SMART; SM00239; C2; 2.  
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.  
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.  
 KW Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;  
 FT Repeat; Alternative splicing.  
 FT NON TER 1  
 FT DOMAIN 400 465 10 X 10 AA TANDEM APPROXIMATE REPEATS OF

FT ZN\_FING 499 523  
 FT ZN\_FING 969 992  
 FT NON\_CONS 1010 1011  
 FT DOMAIN 2300 2325  
 FT DOMAIN 4391 4442  
 FT DOMAIN 4544 4633  
 FT DOMAIN 5031 5121  
 FT VARSPLIC 4404 4404  
 FT  
 FT VARSPLIC 4534 4534  
 FT VARSPLIC 4576 4576  
 FT VARSPLIC 4757 4761  
 FT VARSPLIC 4762 5147  
 FT SEQUENCE 5147 AA; 563537 MW; CDS084990498CD3C CRC64;  
 SQ  
 Query Match 3.1%; Score 161; DB 1; Length 5147;  
 Best Local Similarity 18.7%; Pred. No. 0.61;  
 Matches 114; Conservative 85; Mismatches 189; Indels 222; Gaps 25;  
 QY 91 AESSEKNNNTRIDSMSSVOKD--NFYPHKVKEKLVNVPQNLNLDKSP---TEKSSOYLNQ 145  
 Db 1359 SEEEIKESQEEKDTFKKXDSQQDIPSKDKH-EKSEFVDDITTRREFYDVSVEESESSEN- 1416  
 QY 146 QQTASVCKWQNEGHAEOLLASEPPAGTFLPKOLSNANIGQSPHTDDHSDTDHEDDRNQ 205  
 Db 1417 -----SPVQKRTSVGS-----SSDEYKQEDSQ 1442  
 QY 206 QELTPIKLANTKPTVGDQARSNCCKSGSROSVKCTGCOQEEVDVLPESPLSDVGARDI 265  
 Db 1443 -----GSG-----EEDDFI-RKQIIEASDED 1463  
 QY 266 GTGPKNDNKLGTQSSSLGDSPPFKESPE-----SPVDVNSRNSC 307  
 Db 1464 AGSGDEDFIRNQLKEISSSTESQKKEKTKGKITAGKHRLTKSSTSDDEDAGRHS 1523  
 QY 308 QDSEADE--EFSVPFDEQDDRSSQT-----ANKLSSQAREADGDLRKRYLTKGSEVR 358  
 Db 1524 WHDEDEAPDESPELKRYETKSQSEELVVTGGGLRFRFKIILNSTIADKYSAESSQKK 1583  
 QY 359 LHFQPEGE-----NNAGTSDINAK--PSGNS-----SSINVECRSKQHG 396  
 Db 1584 TSLYDEPEPELEMESLTDSPEDRSRSGESSLHASSFTPGTSPTSVSSIDEDSDSPSHK 1643  
 QY 397 KDSK-----ITD-----HFMFKSKEDRR 416  
 Db 1644 KGESQQRKARHPHGLPTIEDSSSEELREELKEQKREIEQQQKSSKSKK 1703  
 QY 417 KEQCEVRHORTKIPKYPNLPPEKKWLGPIIEMRMKMP-----CGIHLPSLRP 468  
 Db 1704 KDKELRAQR--RRERPKTPPSNLSIED--ASPTELRQAAEWEELHRSSCEYSPSIES 1760  
 QY 469 SASHTVTVRVLLRAGEVPKPTTHYKOLMD-NKHVKMPCSEONLYPVEDENGERTAGSR 527  
 Db 1761 D-----PEGFEISPEKIEIVQVKVKKLP-TAVSLYSPTDEQ----- 1794  
 QY 528 WELIQTALLNKTRPQNLKDALKNVAYSKKWDFTAL--VDFWQKVLSEAE-----A 578  
 Db 1795 -----SIMQKGSQKALKSAEMEYEMMHKTHKYKAPPAANERDEVEFEKEPLYGGMLE 1848  
 QY 579 QHLYQSILPD 588  
 Db 1849 DYIYSLVED 1858  
 RESULT 7  
 DSDPP RAT  
 ID\_DSPP\_RAT STANDARD; PRT; 687 AA.

AC Q62598; P70578; Q9R057;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Dentin sialophosphoprotein precursor [Contains: Dentin phosphoprotein  
GN (Dentin phosphophoryn) (DPP); Dentin sialoprotein (DSP)].  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RN SEQUENCE FROM N.A. (ISOFORM DPP-2).  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=20435277; PubMed=10978503;  
RA Ritchie H.H., Wang L.-H.;  
RT "The presence of multiple rat DSP-PP transcripts.";  
RL Biochim. Biophys. Acta 1493:27-32(2000).  
RN [2]  
RN SEQUENCE OF 1-387 FROM N.A., AND SEQUENCE OF 18-26.  
RC STRAIN=Sprague-Dawley; TISSUE=Odontoblast;  
RX MEDLINE=94148875; PubMed=8106414;  
RA Ritchie H.H., Hou H., Veis A., Butler W.T.;  
RT "Cloning and sequence determination of rat dentin sialoprotein, a  
RL novel dentin protein.";  
RL J. Biol. Chem. 269:3698-3702(1994).  
RN [3]  
RN SEQUENCE OF 421-687 FROM N.A. (ISOFORM DPP-1).  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=9635551; PubMed=8702961;  
RA Ritchie H.H., Wang L.-H.;  
RT "Sequence determination of an extremely acidic rat dentin  
RL phosphoprotein.";  
RL J. Biol. Chem. 271:21695-21698(1996).  
RN [4]  
RN SEQUENCE OF 1-19 FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=99333695; PubMed=10403786;  
RA Yamazaki H., Kuniada T., Miyamoto A., Tagaya H., Hayaishi S.-I.;  
RT "Tooth-specific expression conferred by the regulatory sequences of  
RL rat dentin sialoprotein gene in transgenic mice.";  
RL Biochem. Biophys. Res. Commun. 260:433-440(1999).  
RN [5]  
RN SEQUENCE OF 29-33; 70-79; 93-109; 136-148; 162-188; 266-308; 398-423  
RP AND 426-438, AND PHOSPHORYLATION OF SER-292 AND SER-298.  
RX MEDLINE=21125612; PubMed=11042175;  
RA Qin C., Cook R.G., Orkiszewski R.S., Butler W.T.;  
RT "Identification and characterization of the carboxyl-terminal region  
RL of rat dentin sialoprotein.";  
RL J. Biol. Chem. 276:904-909(2001).  
RN [6]  
RN TISSUE SPECIFICITY.  
RX MEDLINE=98055479; PubMed=9395101;  
RA Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L.,  
RA Hotten D., Papagerakis P., Berdal A., Butler W.T.;  
RT "Dentin sialoprotein (DSP) transcripts: developmentally-sustained  
RT expression in odontoblasts and transient expression in  
RT pre-ameloblasts.";  
RL Eur. J. Oral Sci. 105:405-413(1997).  
CC -!- FUNCTION: DSP may be an important factor in dentinogenesis. DPP  
CC may bind high amount of calcium and facilitate initial  
CC mineralization of dentin matrix collagen as well as regulate the  
CC size and shape of the crystals.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Names=DPP-1; Synonyms=PP240;  
CC IsoId=Q62598-1; Sequence=Displayed;  
CC Names=DPP-2; Synonyms=PP171;  
CC IsoId=Q62598-2; Sequence=VSP\_003855;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Specifically expressed in teeth, mainly in  
CC odontoblasts and transiently in pre-ameloblasts.

CC -!- PTM: DSP is glycosylated.  
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a  
CC frameshift in position 380.  
CC  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF247187; AAK96895.1; -  
CC EMBL; U02074; AAI18932.1; ALT\_FRAME.  
CC EMBL; U63111; AAC52774.1; -  
CC EMBL; AF114987; AAD48588.1; ALT\_TERM.  
CC Biomineralization; Extracellular matrix; Signal; Glycoprotein;  
CC Phosphorylation; Sialic acid; Alternative splicing.  
CC SIGNAL 1 17  
CC CHAIN 18 687 DENTIN SIALOPHOSPHOPROTEIN.  
CC CHAIN 18 447 DENTIN SIALOPROTEIN.  
CC CHAIN 448 687 DENTIN PHOSPHOPROTEIN.  
CC MOD\_RES 57 57 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 226 226 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 253 253 PHOSPHORYLATION (BY CK1) (POTENTIAL).  
CC MOD\_RES 278 278 PHOSPHORYLATION (BY CK1) (POTENTIAL).  
CC MOD\_RES 292 292 PHOSPHORYLATION (BY CK2).  
CC MOD\_RES 298 298 PHOSPHORYLATION (BY CK1).  
CC MOD\_RES 315 315 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 319 319 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 329 329 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 337 337 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 345 345 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
CC MOD\_RES 366 366 PHOSPHORYLATION (BY CK1) (POTENTIAL).  
CC CARBOHYD 55 55 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 189 189 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 312 312 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 369 369 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC VARSPLIC 567 635 Missing (in isoform DPP-2).  
CC  
CC CONFLICT 74 74 N -> D (IN REF. 5; AA SEQUENCE).  
CC CONFLICT 564 564 S -> T (IN REF. 3).  
CC SEQUENCE 687 AA; 70179 MW; 9A845ED6AA31B63 CRC64;  
SQ

Query Match 3.1%; Score 159; DB 1; Length 687;  
Best Local Similarity 20.3%; Pred. No. 0.056;  
Matches 85; Conservative 61; Mismatches 195; Indels 78; Gaps 13;

QY 16 GAAGTSAPTASDSRSPFGRRVLDPKADPVPFRVPPSPACVSGRA-----GPHRG 67  
DB 309 GGDNTSSSEETGIEEGDGTQTTQDNQLSPTEGGIISQAEACPSGQSQNGLETSSTG 368  
QY 68 NATSFVKQKTTTWMDTKG-----PKTAESE-----SKENN---TRIDS 105  
DB 369 NKSITKESGLSGKSDNSNGHEGMDKENSFKQGESDKPQGAEEKSDTHNNMGHGRIGS 428  
QY 106 MMSSVQKDNFYPHKVKLENVQLNLDKSPTEKSSQYLNQQTASVCKWQNEKHARQLL 165  
DB 429 SNSDGHDS-YDFDDSMGDDPNDSNDSNGSDGSDANSESAI-----ENGNHGASY 481  
QY 166 ASEPPAGTPLPKQLGNANIGQSPHTDDSDTDHEEDRNQQLFTPIKLANTKFTVGDGQA 225  
DB 482 TSDSESSDNGSD---SDSHAGEDDSSDDTSDTDDSDNGDDDESCKD-----DESD 529  
QY 226 RNCKCGSRQSVKCTGCGQEEVDVLPSPISDVGAEDIGTPKMDNKLKTGOESLIGDS 285  
DB 530 NSNHNDSDSEKSDS 588  
QY 286 PFPEKESESPMDVNSNSQDSEADETSPVDEQDRSQTANKLSSCOAREADGD 345  
DB 589 SNSSDTSDDS 645



CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; Y19187; CAB60725.1; -  
 CC HSP; P04410; IAA25.  
 CC GO; GO:0045202; C:synaptic junction; ISS.  
 CC GO; GO:0005509; F:calcium ion binding; ISS.  
 CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.  
 CC GO; GO:0005522; F:profilin binding; ISS.  
 CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
 CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.  
 CC InterPro; IPR000008; C2.  
 CC InterPro; IPR001478; PDZ.  
 CC InterPro; IPR001565; Synaptotagmin.  
 CC InterPro; IPR008899; Znf\_picollo.  
 CC Pfam; PF00168; C2; 2.  
 CC Pfam; PF05595; PDZ; 1.  
 CC Pfam; PF05715; Zf\_picollo; 2.  
 CC PRINTS; PR00399; SYNAPTOTAGMN.  
 CC SMART; SM00239; C2; 2.  
 CC SMART; SM00228; PDZ; 1.  
 CC PROSITE; PS00499; C2 DOMAIN 1; 1.  
 CC PROSITE; PS00004; C2 DOMAIN 2; 2.  
 CC PROSITE; PS0106; PDZ; 1.  
 CC Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;  
 KW Repeat.  
 KW NON\_TER  
 FT 1 1  
 FT DOMAIN 258 357  
 FT FT  
 FT ZN\_FING 368 392  
 FT ZN\_FING 836 859  
 FT DOMAIN 2324 2343  
 FT DOMAIN 4414 4493  
 FT DOMAIN 4627 4726  
 FT DOMAIN 5003 5094  
 FT SEQUENCE 5120 AA; A658D9891B65B412 CRC64;  
 Query Match 3.0%; Score 152.5; DB 1; Length 5120;  
 Best Local Similarity 17.9%; Fred. No. 1.9; Indels 151; Gaps 29;  
 Matches 115; Conservative 122; Mismatches 253  
 CC 40 DPKDAPVQFRVPPSPACVGRAGPHRGNATSFVFKQKTIITWDTKPGKTAESKKN 99  
 Db 1302 DEKULPRE---PQKOTISOESPPSPDLAKL---ESTVLISLEAQAATLTDKSVK 1354  
 Qy 100 NTRIDSMMSVQKNFYHVKLENVQ-Q-INLQKSPTEKSSQVYNQOQTASVCKW--QN 156  
 Db 1355 --ELYETYSQTKDQ---HKTPLPVTPESSYSSDEEDLEAIQE--GERTIAADSKG 1407  
 Qy 157 EGKHAQQLLASEPPA-----GTPLPKQLSNANIQOSPHTDHSDTDHEED 201  
 Db 1408 QTDYKEEDGGNDTARRQRYDSVEDSESENSPVPRKRRASVGS--SSDYKRDSDGS 1466  
 Qy 202 RDNOQFLTPIKLANTKPTVGQARSNCKSGRSQSVKCTGCOQEEVDVLPESPLSDVG 261  
 Db 1467 GDEEDFI-----RKQIEMSADEADSGEDDEFIRN---QLKEISVTSQKKEVK 1514  
 Qy 262 AEDTGTGPNKLTGQSSSLG-----DSPPEKSEPE-----SPMDVD-- 301  
 Db 1515 SKAKGTGK--HRRMARKSSAGYDEAGRRHSWHDDDDTETDEGPEPKYRTKSDQGBEL 1572  
 Qy 302 -----NSRNSCODSEADETSPV--FDEQDRSSOTANKLSSQAR--EA 342  
 Db 1573 AISGGGLRRFKTIENLTITSKSETPEQQKGLYFDEEPELEMSLTDSPEERSGEG 1632  
 Qy 343 DGDLRKRYLTGSEVRLHFQFEGENNA-----GTSDLNAK-----PSGNSSS 384  
 Db 1633 SSSLHASFPTGCTSTSVSLDESDSSPHKLGESKQQRKARHRSGHPLLTIEDSS 1692  
 Qy 385 LNVECRSKQHGKRSK--ITDFMRISKSDRRKEQCEVHRHQRTERIKPIKYPNLPPE 442  
 Db 1693 EBEELREBEELKEQEKQRELEQQQKSSSKSKKDKDELRAQR--RRRPRKTPPSNLSPI 1751

Qy 443 KKWLGTPIEMRKMPR-----CGIHLPSLRPSASHTVTVRVDLIRAGEVPEKPFPTHY 494  
 Db 1752 ED--ASPTBELRQAEMELHRSSCSEYSPSIESDPEGEISPEKIEVKVYK-LFTAV 1808  
 Qy 495 KDLMDNKHVKMPCSEQNLYPVEDENGERTAGSRWELIQTALLNKFTTRPQNLKDALIKYV 554  
 Db 1809 S-----LYSPDTEKLIGALKEESQKTLKSAEYVE-EMIKHTKSKS-----FQI 1853  
 Qy 555 AYSKKWDFVALVDFWPKVLEEAE-----AQHLYQSILPD 588  
 Db 1854 ASEK-----DEVFEKESLYGGLIEDIYIESLIED 1883  
 RESULT 15  
 ATRX CAEEL  
 ID ATRX CAEEL STANDARD; PRT; 1359 AA.  
 AC Q9U7EO; C02061;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).  
 GN XNP-1 OR B0041.7.  
 OS Caenorhabditis elegans.  
 OC Rukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=99365296; PubMed=10433961;  
 RA Villard L., Fontes M., Ewbank J.J.;  
 RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to  
 RT the human XNP/ATR-X gene.";  
 RL Gene 236:13-19(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Fulton R., Wohldmann P.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Could be a global transcriptional regulator. Modifies  
 CC gene expression by affecting chromatin (Potential).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF134186; AAD55361.1; -  
 CC EMBL; AF000196; AAC24256.1; -  
 CC PIR; T34036; T34036.  
 CC WormPep; B0041.7; CE17314.  
 CC InterPro; IPR001410; DEAD.  
 CC InterPro; IPR001650; Helicase C.  
 CC InterPro; IPR000330; SNF2\_N.  
 CC Pfam; PF00271; helicase\_C; 1.  
 CC Pfam; PF00176; SNF2\_N; 1.  
 CC SMART; SM00487; DEXDC; 1.  
 CC SMART; SM00490; HELIC; 1.  
 CC PROSITE; PS00690; DEAD ATP HELICASE; Nuclear protein; ATP-binding;  
 KW DNA repair; Hydrolase; Helicase; Helicase; Nuclear protein; ATP-binding;  
 KW DNA-binding.  
 FT NP\_BIND 496 503  
 FT SITE 636 639  
 FT DEAH\_BOX  
 FT DOMAIN 67 70  
 FT POLY-ASP.  
 FT DOMAIN 266 272  
 FT POLY-GLU.  
 FT DOMAIN 276 281  
 FT POLY-LYS.  
 FT DOMAIN 372 375  
 FT POLY-LYS.  
 FT DOMAIN 603 608  
 FT POLY-LYS.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 17:10:29 ; Search time 169.229 Seconds

(without alignments)

13416.302 Million cell updates/sec

Title: US-09-302-812-1

Perfect score: 7242

Sequence: 1 accggaaggaacgaagcc.....aaatttcattcaaaaaa 4070

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Xgapext 0.5	
Xgapop 6.0 , Xgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 2298626

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/633148/runat.26052004.150056.5800/app.query.fasta.1.12437  
-DB=Published Applications AA -QFW=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=633148 @CGN 1 1 79 @runat.26052004.150056.5800  
-NCPU=6 -ICPU=3 -NO WRAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FCGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2:	/cgn2_6/ptodata/1/pubpaa/PTC_NEW_PUB.pcp.*
3:	/cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pcp.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6:	/cgn2_6/ptodata/1/pubpaa/PTCUS_PUBCOMB.pcp.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

1	5184	71.6	977	9	US-09-973-451-2	Sequence 2, Appli
2	4635.5	64.0	976	9	US-09-973-451-4	Sequence 4, Appli
3	4332.5	59.8	968	9	US-09-973-451-6	Sequence 6, Appli
4	1063.5	14.7	768	9	US-09-973-451-8	Sequence 8, Appli
5	594	8.2	546	12	US-10-425-114-60000	Sequence 60000, A
6	465	6.4	726	9	US-09-973-451-10	Sequence 10, Appli
7	205.5	2.8	200	12	US-10-424-599-184988	Sequence 184988,
8	184	2.5	180	12	US-10-424-599-156445	Sequence 156445,
9	184	2.5	2665	9	US-09-864-761-34248	Sequence 34248, A
10	184	2.5	3664	12	US-10-263-929-143	Sequence 143, App
11	184	2.5	3664	14	US-10-177-293-423	Sequence 423, App
12	175.5	2.4	1163	12	US-10-221-278-197	Sequence 197, App
13	175.5	2.4	1163	15	US-10-291-172-197	Sequence 197, App
14	174.5	2.4	1462	14	US-10-287-218-17	Sequence 17, Appli
15	173.5	2.4	1163	10	US-09-932-257A-3	Sequence 3, Appli
16	173	2.4	1210	14	US-10-205-823-264	Sequence 264, App
17	172.5	2.4	2828	12	US-10-072-012-607	Sequence 607, App
18	170.5	2.4	2828	9	US-09-905-129-21	Sequence 21, Appli
19	170.5	2.4	2828	9	US-09-991-630-21	Sequence 608, App
20	170.5	2.4	2828	12	US-10-072-012-608	Sequence 21, Appli
21	170.5	2.4	2828	12	US-10-454-351-21	Sequence 54, Appli
22	170.5	2.4	2828	12	US-09-802-318-21	Sequence 110, App
23	170.5	2.4	2828	14	US-10-176-847-54	Sequence 49, Appli
24	170.5	2.4	2828	14	US-10-301-822-49	Sequence 126, App
25	170.5	2.4	2828	14	US-10-032-189-126	Sequence 58, Appli
26	170.5	2.4	2828	15	US-10-295-027-58	Sequence 1175, Ap
27	170.5	2.4	2828	15	US-10-295-027-1175	Sequence 5013, Ap
28	170.5	2.4	2828	15	US-10-369-493-5013	Sequence 4, Appli
29	170.5	2.4	6642	15	US-10-117-604-4	Sequence 11, Appli
30	169.5	2.3	913	13	US-09-973-451-11	Sequence 2651, Ap
31	167	2.3	31	9	US-09-973-451-11	Sequence 233915,
32	165.5	2.3	2515	12	US-10-276-774-2651	Sequence 2, Appli
33	164.5	2.3	300	12	US-10-424-599-233915	Sequence 224, App
34	164.5	2.3	1804	14	US-10-275-140-2	Sequence 67803, A
35	164	2.3	1390	12	US-10-092-900A-224	Sequence 72939, A
36	163	2.2	747	12	US-10-425-114-72939	Sequence 43444, A
37	162.5	2.2	1109	12	US-10-425-114-72939	Sequence 1, Appli
38	162	2.2	748	9	US-09-864-761-43244	Sequence 142, App
39	162	2.2	1404	9	US-09-811-045A-1	Sequence 4345, Ap
40	161.5	2.2	5560	12	US-10-263-929-142	Sequence 8, Appli
41	161	2.2	691	15	US-10-108-260A-4345	Sequence 4, Appli
42	161	2.2	951	14	US-10-379-616-8	Sequence 196154,
43	161	2.2	1420	14	US-10-379-616-8	Sequence 3747, Ap
44	160	2.2	498	12	US-10-424-599-196154	
45	160	2.2	748	15	US-10-108-260A-3747	

ALIGNMENTS

RESULT 1

US-09-973-451-2

Sequence 2, Application US/09973451

Patent No. US20020132328A1

GENERAL INFORMATION:

APPLICANT: JACOBSON, Myron K.

APPLICANT: JACOBSON, Elaine L.

APPLICANT: AM, Jean-Christophe

APPLICANT: LIN, Winston

TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE

TITLE OF INVENTION: (PARG) ENZYMES,

TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

TITLE OF INVENTION: THEREWITH

FILE REFERENCE: NIAD 201

CURRENT APPLICATION NUMBER: US/09/973,451

CURRENT FILING DATE: 2001-10-09

PRIOR APPLICATION NUMBER: US/09/302,812

PRIOR FILING DATE: 1999-04-30

PRIOR APPLICATION NUMBER: 60/083,768

PRIOR FILING DATE: 1998-05-01

NUMBER OF SEQ ID NOS: 38

SEQ ID NO 2

LENGTH: 977

TYPE: PRT

```

; ORGANISM: Bos taurus
; FEATURE:
US-09-973-451-2

Alignment Scores:
Pred. No.: 0 Length: 977
Score: 5184.00 Matches: 977
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 71.58% Indels: 0
DB: 9 Gaps: 0

US-09-302-812-1 (1-4070) x US-09-973-451-2 (1-977)
QY 258 ATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGAGCGCGCTGCA 317
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20
QY 318 ACTTCTCCGCGCGCGCTCGAGCGCGCGAGCTTCCCGCGGAGGAGCGCGCTCCTC 377
Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40
QY 378 GATTCCAGGACGCTCGGTGCGATTGAGGTCGCGCTCCGCGCTCGTCAGGCTCGCGCTG 437
Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60
QY 438 GCGCGCGCGGACAGACCGGAGCGCGCGCCACCTCTCTGTTTTCACAAAGAACTATA 497
Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80
QY 498 ACCAGTTGGATGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAGTTGTCATAGTAA 557
Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100
QY 558 GAAACCAACATACAGAGAGAGATCATGATGATGCTGTACAAAGAAATACTTTAT 617
Db 101 GluAsnAsnAsnThrArgGluGluSerMetSerSerValGlnLysAspAsnPheTyr 120
QY 618 CAACATAACATGAAATAATAGAAAATGTTTCTCAGCTAGGTTTGTATAGTCACCAATT 677
Db 121 GlnHisMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140
QY 678 GAAAGAGTACAGTATTTGAGCAGCAGCAGCAGCTCGGCTATGCTTAAGTGCAGAAAT 737
Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysLysTrpGlnAsn 160
QY 738 GAAGGCGCACACTCAGACGCGCTTTGGAAGTGAACCTCCAGCGGTAACTCTGTGTACCA 797
Db 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180
QY 798 GAGCAGTTCAGTAAATGCTAATGCTCAGTCAGTCGTCGCCAAAGGATGATCAGTGACACA 857
Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200
QY 858 AATAGTACGAGAGTACAGATAATCAGCAGTTTTTGGACATGATAAGCTTGGCAATGCA 917
Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
QY 918 AAGCAGACGATGGAAGATGAACAGGCGGACAGAGCCAGGAGCCAGGAGTGTGGCAAG 977
Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240
QY 978 GCTTGCCATCTCGCAGAGCGCTGTGACGGGTGTGACAGAGGAGAGACAGCTGGTGTCC 1037
Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260
QY 1038 GAGAGCCCTTGTGGACACTGGCTCTCAGGATGTTGTTGATGCTGACATGAAAATGCCAAC 1097
Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280
QY 1098 AGATTGAATAGACAGAAGTAGTCTAGGAATCTCCTCCATTGAGAAAAGAAAGTCAA 1157
Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300

```

```

QY 1158 CTGAGTCACCAATGATGTAGATAATTTCCAAAATAGTTGTGTCAGATTTCAGAGCAGAT 1217
Db 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320
QY 1218 GAAGAGACAAAGTCCAGGTTTTTCATCAACAGAGAAAGATAGCAGTTCTGCTCAACAGCAAAAT 1277
Db 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACATGAGTTGAGGAACGGTCTCTGCT 1337
Db 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360
QY 1338 AAGGAGGTGAGATTCGATTATCATTTCCAAATTTGAGGAGGAGAGAGTGCAGTGAATG 1397
Db 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380
QY 1398 AATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCCTGAAATGATGAGTGCAGAAAT 1457
Db 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400
QY 1458 TCTAAGCAACATGGGAGAAAGGATTCTTAAATTCACAGATCATTTTCATGAGTGCACAAA 1517
Db 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420
QY 1518 GCAGAGACAAAGAAAGAAACAATGTGAAATCAACATCAAAAGAACACAGAAAGAAAGATC 1577
Db 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440
QY 1578 CCTAATAATACATTCACCTCTCCAGATTAAGAAATGGCTTCGAACTCTATGAG 1637
Db 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 460
QY 1638 GAGATGAGAGAAATGCAAGGTGTGGATCCGGTCCGCTCCCTTGAGACATCTGCCAAT 1697
Db 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480
QY 1698 CACACAGTCATCTCGGTGATCTTTGCGAATAGGAGAGTCTCTAAACCTTTCCCA 1757
Db 481 HisThrValThrIleArgValAspLeuLeuArgLysGlyGluValProLysProPhePro 500
QY 1758 ACACATTTAAAGATTGTGGACAAACAGCATGTTAAGATGCTTGTTCAGAACAAAAC 1817
Db 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520
QY 1818 TTGTACCTGTGGAAGATGAGAAATGTGACGAGCTGCAGGACGCGCTGGGAATCATT 1877
Db 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540
QY 1878 CAGACTGCATCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAAGGATGCTATTCTGAAG 1937
Db 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560
QY 1938 TACAATGTGSCATATTCTAAGAAATGGATTTCAGCTTTGATTGATTTCGGGATAG 1997
Db 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580
QY 1998 GTACTAGAAGACAGACAGCTCAACCTTGTATCAGTCCATCTTGTGCTGATATGTTGAAA 2057
Db 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600
QY 2058 ATTGCATCTGTCTGCCAAATATTGTACCCAGCAATACCACTCTCTGAAACAGAGATG 2117
Db 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620
QY 2118 AATCATTTCCATCAATGTGCAGGAAACAGATTGCGAGTCTTTTAGTAATGTTCTTCTTC 2177
Db 621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640
QY 2178 TGCAGGTTTCCACGACCGCAATGCGAGATCAATACAGATATTCCAGTATTCCAGATATT 2237
Db 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660

```





QY 918 AACGACGACGATGAAGATGAACAGGCGCAGAGAGCCAGAGACCCACAGAGTGTGGCAAG 977  
DB 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
QY 978 GCTTGGCATCTCGCAGAGCTGTGCGAGGTGTCTCAGCAGAGGAGCAGACGCTGGTGTCC 1037  
DB 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuValPro 259  
QY 1038 GAGAGCCCTTGTGGACACTGGCTCTGAGAGTGTGGTACTGCTGAAATGCAAC 1097  
DB 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
QY 1098 AGATTGAATAGACAAGAAAGTAGTCTAGGAAATCTCTCCATTGAGAAAGAAAGTCAA 1157  
DB 280 LysLeuLeuArgGlnGluSerCysLysGluAsnSerProProGluLysGluSerGlu 299  
QY 1158 CTTGAGTCCCAATGGATGTAGATAATTCAAAATAAGTGTCTCAGGTTTCTCAGAGCAGAT 1217  
DB 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
QY 1218 GAAGAGCAAGTCCAGGTTTTGATGAACAGAGATAGCAGTCTGCTCAACAGCAAAAT 1277  
DB 320 GluGluThrSerProGlyPheAspGluGluLysAspGlySerSerSerGlnThrAlaAsn 339  
QY 1278 AAACCTTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGGAGCGGTCTCTGCT 1337  
DB 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
QY 1338 AAGGAGGTGAGATTCGATTACATTTCCAAATTTCAAGCAGAGAGAGTCTCAGCTGGAATG 1397  
DB 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
QY 1398 AATGATGTGAATGCAACAGCTGGAAGTACTTCTAGCTGAATGAGAGTGCAGAAAT 1457  
DB 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCAAAATCAAGATCATTTTCATGAGAGTGCACAAA 1517  
DB 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
QY 1518 GCAGAGACAAGAAAGAAACAATGTGAATGAACATCAAGACAGAGAGAGATC 1577  
DB 420 AlaGluAspArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
QY 1578 CCTAAATACATTCACCTTCTCCAGATAAGAAATGGCTTGGAACTCTTATTGAG 1637  
DB 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
QY 1638 GAGATGAGGAGAAATGCCAAGGTGGGATCCGGCTGCTCCTCTGAGACCATCTGCCAAT 1697  
DB 460 GluMetArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
QY 1698 CACACGTGACTATTCGGGTAGATCTTTTGGAAATAGAGAGTCTTAAACCTTTCCCA 1757  
DB 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
QY 1758 ACATATTTAAAGATTTGTGGCAACAAGCATGTTAAGATGCTTGTTCAGAACAAAC 1817  
DB 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
QY 1818 TTGTACCTGTGGAAGATGAATGTGTGAGGAGCTGCAGCAGCCGGTGGGAATCAATT 1877  
DB 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
QY 1878 CAGACTGCATCTTCAACAGCTCACTCGGCCAGAACCTTGAAGGATGCTATTCTGAAG 1937  
DB 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559  
QY 1938 TACAATGTGGCATATTTCTAAGAAATGGGACTTTACAGCTTTGATTGATTTCTGGGATAAG 1997  
DB 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
QY 1998 GTACTAGAAGACGAGAGCTCAACACTGTGTATCAGTCCATCTTGGCTCGATATGGTGAA 2057

DB 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
QY 2058 ATTGCACTCTGTCTGCCAAATATTGTACCCAGCAATACCCTCTCTGAAAACAGAGATG 2117  
DB 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
QY 2118 AATCATTTCCATCACATGTCTCAGAGAACAGATTGCCAGTCTTTTACCTAATGTTTCTTC 2177  
DB 620 AsnHisSerIleThrMetSerGlnGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
QY 2178 TGCACGTTTCCAGCAGCGAATGCCAAGATGAATCAGAGTATTCAGATTATCCAGATATT 2237  
DB 640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
QY 2238 AACTTCAATCGGTGTGTGAAGGACGTTTCATCAAGGAAACCCAGAGAGCTTAAACGCTC 2297  
DB 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
QY 2298 TTCTGCTACTTTAGAAAGAGTCCAGAGAAAAACCCACTGGGTGTGTGACATTTCAAGA 2357  
DB 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
QY 2358 CAGAGTCTTGAGATTTTCCAGAGTGGGAAAGATCTGAAAACCTCTGACTCGACTGCAT 2417  
DB 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
QY 2418 GTCACTTACGAAGGTACCATAGAGAAACCGCCAGCGCATGTCTACAGTGGATTTTGA 2477  
DB 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
QY 2478 AACCGTTTCTGAGGTGTGTAAACAGTGCAGGACTGTGTGCAAGAAAAATCCCGTTT 2537  
DB 740 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
QY 2538 TTAAATCAACCTCAGCTGATTGTTTCCAGCTCTTCTCAGTGTGTGAGTCAACAATGAA 2597  
DB 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspPheAsnGlu 779  
QY 2598 TGTCTTATCATCACAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
DB 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
QY 2658 CGCTGGCCCGGAGCCATGAGACAGAGCGAAAGGACGACTGGCAGCGCGCAGCT 2717  
DB 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspCysGluArgArgCysThr 819  
QY 2718 GAGATCGTCCCATCCAGCGCTCCACTTCAGACGCTACCTCGACCATGTTGTGCCCGAG 2777  
DB 820 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 839  
QY 2778 AAGATCAGCGGAGCTTAAACAGGCTTACTGTGGATTTCTTCGCTCGAGTTTCTTCA 2837  
DB 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
QY 2838 GAGAACCTGTCTCAGTGGCTACAGGAACTGGGCTGTGTGCTTGGGGGTGATGCT 2897  
DB 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
QY 2898 AGACTAAAGCCCTTAATACAGATCCCTGGCAGCTGTGTAGCTGAGCAGACGCTGGTTAT 2957  
DB 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
QY 2958 TTCACCTTTGGGAGCTCAGAACTGATGAGAGACATTTACAGCATGTATATCTCTCACT 3017  
DB 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919  
QY 3018 GAGAGGAACTGACTGTTGGAGAGATATTAAGCTGCTAGCATATTAACAATGAGAA 3077  
DB 920 GluArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 939  
QY 3078 TGCAAGAACTGCTCCACCCCGGACAGACATCAAGCTTTTATCCATTTCATATACCATGCA 3137

```

Db      940 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 959
QY      3138 GTTCAGTCTCTGACACAGACACCAACACCGCGGACAAAGAGCGGG 3185
Db      960 ValGluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGly 975

RESULT 3
US-09-973-451-6
; Sequence 6, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES,
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-973-451-6

Alignment Scores:
Pred. No.: 3,826-304 Length: 968
Score: 4332.50 Matches: 820
Percent Similarity: 89.86% Conservative: 57
Best Local Similarity: 84.02% Mismatches: 90
Query Match: 59.82% Indels: 9
DB: Gaps: 5

US-09-302-812-1 (1-4070) x US-09-973-451-6 (1-968)
QY      258 ATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGGACGCGCTGCA 317
Db      1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19
QY      318 ACTTCTCCGCGCGCGCTCGGACGCGCGGAGCTTCCCGCGGACGAGCGCGCGCTCCTC 377
Db      20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39
QY      378 GATTCCAGAGACGCTCCGGTGCAGTTCAGGGTCCCGCGCTCCTCGTCAGGCTCGCGCTG 437
Db      40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
QY      438 GCGCGCGCGGACAGACCCGAGGACGCGCACCTCTCTCTTTTCAACAGAGAGACTATA 497
Db      60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79
QY      498 ACCAGTTGGATGGACACTAAAGGAATCAAGACAGTTCGAATCAGAAAGTTTGCATAGTAA 557
Db      80 ThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96
QY      558 GAAACACAAATACAGAGAGAAATCCATGATGATGTTCTGTACAAAAGATACTTTAT 617
Db      97 GluAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 116
QY      618 CAACATAACATGGAATAATTAGAAAATGTTTCTCAGCTAGTGTGTTGATAGTACACCAATT 677
Db      117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136
QY      678 GAAAAAGGTACACAGTATTTTGAAGCAGCATCAGAGCTCGGGCTATGTGTGAAGTGCAGAA 737

```

```

Db      137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156
QY      738 GAAGGGCCACACTCAGAACGGCTTTTGAAGTCAACCTCCAGCGGTAACTCTGTGTACCA 797
Db      157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProAlaGlyThrProLeuPro 176
QY      798 GAGCAGTTTCAGTAATGCTGATCGATCGATCGTGTCTCCCAAGAGGATGATCAGAGTGCACA 857
Db      177 LysGlnLeuSerAsnAlaAsnIleGlnSerProHisThrAspAspHisSerAspThr 196
QY      858 AATAGTGAAGAGAGTAGAGATAATCAGCAGTTTTCACACATGTAAAGCTTGCAGTACCA 917
Db      197 AspHisGluGluAspArgAspAsnGlnPheLeuThrProLysLeuAlaAsnThr 216
QY      918 AAGCAGAGCAGTGAAGATGAACAGCGGACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 977
Db      217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233
QY      978 GCTTGCCATCTCGCAGAACGCTGTGACGGGTGTGACAGAGGAGGACAGACAGTGTGTGTC 1037
Db      234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253
QY      1038 GAGAGCCCTTGTGCGACACTGGCTCTCAGCATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
Db      254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273
QY      1098 AGATTGAATAGACAGAAAGTAGTCTAGGAATCTCTCCATTTTGAGAAAGAAAGTGA 1157
Db      274 LysLeuThrGlyGlnGlnSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293
QY      1158 CTTGAGTCAACCAATGATGATTAATTCACCAAGAGATAGCAGTTCCTCAACAGCAAAAT 1277
Db      294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313
QY      1218 GAAGAGACAGTCCAGTCTTTCATCAACAGAGAGATAGCAGTTCCTCAACAGCAAAAT 1277
Db      314 GluGluThrSerProValPheAspGluGlnAspAspArgSerSer---GlnThrAlaAsn 332
QY      1278 AAACCTTCAAGGTTCCAAAGAGAGCTGACACTGAGTTCAGGAGAGCGGCTCTGTCTGT 1337
Db      333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352
QY      1338 AAGGAGAGTGCAGATTTCATTTCAATTTGAAGAGAGAGAGAGTGCAGTGCAGTGAATG 1397
Db      353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaAlaGlyThr 371
QY      1398 AATGATGTGAATGCCAAACGACCTCGAAGTACTTCTAGCTCAATAGTAGAGTGCAGAAAT 1457
Db      372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerSerLeuAsnValGluCysArgSer 391
QY      1458 TCTAAGCAACATGGGAGAAAGGATTCATAAATCAACAGATCATTTTCATGAGTGCAGCAAT 1517
Db      392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgIleSerLys 411
QY      1518 GCAGAGCAAAAGAAAACAAACAAATGTGAATCAACATCAAGAGACAGAGAGAGAGATC 1577
Db      412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431
QY      1578 CCTAATATACATTCACCTCCTCCTTCTCCAGATTAAGAAATGGCTTGGAACTCCTATTAG 1637
Db      432 ProLysTrpIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProLysGlu 451
QY      1638 GAGATCAGGAGATGCAAGAGTGTGGGATCCGGCTCCCTCCCTTGTGACCATCTGCCAAT 1697
Db      452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471
QY      1698 CACACAGTACTATTCGGGTAGATCTTTTCGGAATAGGAGAGTTCCTTAACCTTCCCA 1757
Db      472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491
QY      1758 ACACATTTTAAAGATTTGTGGGACACACAGCATGTTAAGATGCTTCTTTCAGAACAAAC 1817

```

Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 QY 1818 TTGTACCTGTGGAGATGAGATGTGTGGAGCTGCGAGCGCGGTGGGAATCATTT 1877  
 Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
 QY 1878 CAGACTGCACTTCTCAGAGCTCACTCGGCCCCAGAACCTGAAGATGCTATTCTGAAG 1937  
 Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
 QY 1938 TACATGTGCGCATATTCTAAGAAATGGACTTTACAGCTTTTGATGATTCTCGGATAAG 1997  
 Db 552 TyrAsnValAlaTyrSerLysLysTyrAspPheThrAlaLeuValAspPheTrpAspLys 571  
 QY 1998 GTACTAGAACAGACAGAGCTCAACTGCTGTATCACTCCATCTGCTGATATGGTGAAA 2057  
 Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
 QY 2058 ATTGCATCTGTCTGCCAATATTGTACCCAGCCCAATACCACTCTCTGAAACAGAAAGATG 2117  
 Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611  
 QY 2118 AATCATTTCCATCAATGTCCAGGACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTC 2177  
 Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 QY 2178 TGACGTTTCCAGACGCAATGCCAGATCAATACAGATTTCCAGTTATCCAGATATT 2237  
 Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerTyrProAspIle 651  
 QY 2238 AACTCAATCGGTGTTTGAAGACGTTTCATCAAGAAACACAGAGACGTTAAAGCGTC 2297  
 Db 652 AspPheAsnArgLeuPheGluArgSerSerArgLysProGluLysLeuLysThrLeu 671  
 QY 2298 TTCTGCTACTTTAGAAGAGTCACAGAGAAAACCCACTGGTGGTGGAATTCACAGA 2357  
 Db 672 PheCysTyrPheArgArgValThrGluLysProThrGlyLeuValThrPheThrArg 691  
 QY 2358 CAGAGTCTGAAGATTTCCAGAGTGGGAAGATGTGAAAACCTCGACTCGACTGCAT 2417  
 Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
 QY 2418 GTCACTTACGAAGTACCATAGAGAAACCGCCAGGCGCTGTACAGTGGGATTTGCA 2477  
 Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
 QY 2478 AACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
 Db 732 AsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluIleArgPhe 751  
 QY 2538 TTAATCAACCTGATTTGATTTTCAGGCTCTTCACTGAGTGTGGATCCACAATGAA 2597  
 Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
 QY 2598 TGTCTTATCATCAGCTACTGACAGTACAGTACAGTACAGTACAGTATCGCGAACATAC 2657  
 Db 772 CysLeuIleLeuThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
 QY 2658 CGCTGGGCCCGGACCATGAGACAGAGCGAAGAGGACGACTGGCAGAGCGCGACACT 2717  
 Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCysThr 811  
 QY 2718 GAGATCGTGGCATCGACGCTTCCACTTACAGCTACCTCGACCTGTTGTCGCCGAG 2777  
 Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
 QY 2778 AAGATCAGACGGAGCTTAAACAAGCTTACTGTGATTTCTTCGTTCTGGAGTTTCTTCA 2837  
 Db 832 LysValArgArgLysLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
 QY 2838 GAGAACCTGTCTGCACTGCTGACAGAAACCTGGGCTGTGTGCTTTGGGGGTGATGCT 2897  
 Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871

QY 2898 AGACTAAAGCCCTTAATACAGATCTGCGAGCTGCTAGCTGAGGAGACGCTGGTTAT 2957  
 Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 891  
 QY 2958 TTCACTTTGGGAGCTCAGAACTGATGAGACATTTACAGATGATGATGATGATGAT 3017  
 Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
 QY 3018 GAGAGAAACTGACTGTTGGAGAAGTATATAAGCTGCTGCTACGATATTACCAATGAAGA 3077  
 Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
 QY 3078 TGCAGAAACTGCTCCACCCCGGACAGACATCAAGCTTTATCCATTCATATACCATGCA 3137  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3138 GTTGAAGTCTGTACAGACACACACACACACACACACACACACACACACACACAC 3185  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967  
 RESULT 4  
 US-09-973-451-8  
 ; Sequence 8, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; CURRENT FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-973-451-8  
 Alignment Scores:  
 Pred. No.: 6,12e-68 Length: 768  
 Score: 1063.50 Matches: 244  
 Percent Similarity: 57.31% Conservative: 93  
 Best Local Similarity: 41.50% Mismatches: 198  
 Query Match: 14.69% Indels: 53  
 DB: 9 Gaps: 15  
 US-09-302-812-1 (1-4070) x US-09-973-451-8 (1-768)  
 QY 1506 AGAGTCCCAAGACGAGGACAAAGAAAGAAACAAATGTGAATGAACATCAAGAAACA 1565  
 Db 45 ArgMetSerLysSerProAspGlyGlyIleSerGluIleGluThrGluGluGlu----- 62  
 QY 1566 GAAAGGAAGATCCCTAAATACATTCACCTCACCTC-----TCTCCAGATAAGAAATGG 1619  
 Db 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTip 74  
 QY 1620 CTTTGGAACTCTATTTAGGAGATG-----AGGAAATGCCAAGGTGTGGATCCGCTG 1673  
 Db 75 ArgGlyValSerMetGluAlaIleHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94  
 QY 1674 CTTCCCTTGAGACCATCTCGCCNAATCACACAGTACTATTCCGGGTAGATCTTTTCGGAATA 1733  
 Db |||||:





Db 97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116  
QY 1131 -----TCTCCTCCATTGAGAAAGTGAACCTGGAGTCACCAATGGATGA--- 1178  
Db 117 HisLysSerThrGluProMetGluGluAspValAsnAsnLysSerAsnIleAspValAla 136  
QY 1179 -----GATAATTCCAAAAATAGTTCTCCAGAT 1205  
Db 137 IleAsnSerAspGluAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156  
QY 1206 TCAGAGCAGATGAAGAGACAGTCCAGGTTT-----GATGAACAGAA 1250  
Db 157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuIleGlu 176  
QY 1251 -----GATAGCAGTTCTGCTCAA 1268  
Db 177 TyrProGlyIleMetLysAspThrThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
QY 1269 ACAGCAATAAACCCTTCAAGGTTTCCAAAGAGAGAGCTGACACTGAGTTGAGGAAGCGG 1328  
Db 197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
QY 1329 TCCTCTGTAAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGAGAGAGATCGA 1388  
Db 212 -----ThrPheValGlyGluAspSerLys 219  
QY 1389 GCTGGAATGAATGATGTGAATGCAACAGCCTGGAAGTACTTCTAGCTGATGTAGAG 1448  
Db 220 Ala-----ThrLysThrValArgThrSerSerSerPheLeuSerThr----- 234  
QY 1449 TGCAGAAATTTCTAAGCAACATGGGAGAAAGGATTTCTAAATACAGATCATTTTCATGAGA 1508  
Db 234 ----- 234  
QY 1509 GTGCCAAAGCAGACAAAAGAAAAGACATGTGAATGAAATCAACATCAAGAACAGAA 1568  
Db 235 ValSerThrCysGlnAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
QY 1569 AGAAGATCCCTAAATACATTTCCACCTCACCCTTCT-----CCAGATAAGAAATGGCTT 1622  
Db 255 LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----- 270  
QY 1623 GGAATCTCTATTGAGGAGATGAGAGAAATGCCAAGGTGGGATCGCGTCCCTCTG 1682  
Db 271 -----LeuAsnLysVal 274  
QY 1683 AGACCATCTGCCAATCACACAGTACTATTTCGGGTAGATCTTTTGCAGATGAGAGATT 1742  
Db 275 AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
QY 1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTTGTGGGCAACAAGCAT----- 1790  
Db 286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304  
QY 1791 GTTAGATCCTTTGTCAGAACAAACTGTGACCTGTGGAGATGAGAAATGGTAGCGA 1850  
Db 305 IleValLeuProGlnArgTyrArgGluPhe-----AspSerArgGlyArgArg 320  
QY 1851 GCTGCAGGCGCGGTGGGAATCTCATTCAGATGCACTTCTCAACAGGCTCACTCGGCC 1910  
Db 320 ----- 320  
QY 1911 CAGAACCTGAAGATGCTATTCTGAAGTACAAAT-----GTGGCATATTCTAAGAA 1961  
Db 321 -----ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337  
QY 1962 TGGGACTTTACAGCT-----TTGATTGATTTCTGGAT----- 1994  
Db 338 TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTyrGluPheAsp 357  
QY 1995 ---AAGTACTAGAACAGACAGACTCAACACTTGATCATCTGATCCATCTTGCTGATATG 2051  
Db 2051 ----- 2051

Db 358 ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377  
QY 2052 GTG-----AAAATGTCACCTCTGTCTG 2072  
Db 378 ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaAlaGlyLeuThrAla 397  
QY 2073 CCAATATTATTGTACCCAGCAATACCACTCCTGAAACAGAGATGAATCATTCATCACA 2132  
Db 398 GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416  
QY 2133 ATGTTCAGAGAACAGATTCCTCAGTCTTTTGTAGTAATGCTTTCTTCTGCAGGTTCCACGA 2192  
Db 417 LeuSerHisLysGlnCysAlaAlaLeuValAlaArgMetPhePhe----- 431  
QY 2193 CGCAATGCCAAGTGAATCAGATATCCAGTATTCAGTTATCCAGATATTAACTTCAATCGTTG 2252  
Db 432 -----AlaArgProAspSerProPheSer-----PheCysArgIle 443  
QY 2253 TTTGAAGGAGCTTCATCAAGGAAACAGAGAAAGCTTTAAACCGCTCTTCTGCTACTTTAGA 2312  
Db 444 LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463  
QY 2313 AGAGTCACAGAGAAAACCCACTGGTGGTGACATTC-----ACAAGACAGAGTCTT 2366  
Db 464 LysMetSerMetAspProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483  
QY 2367 GAAGATTTTCCAGATGGGAAAGATGTGAAACACTCCTGACTGCTGCTGATGCTACTTAC 2426  
Db 484 AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503  
QY 2427 -----GAAGGTACCATAGAAGAAACCGCCAGGCGATCTACAGGTGGATTTTCAAAAC 2480  
Db 504 PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522  
QY 2481 CGTTTCGTGGAGTGGTGTAACTGACAGTGTGACAGACTTGTGCAAGAGAAATCGCTTTTA 2540  
Db 523 GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluIleArgPheLeu 542  
QY 2541 ATCAACCTCGATGATGTTTTCACGGCTCTTTCACGTAGTCTGATGATCACAATGAT 2600  
Db 543 MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuAla 562  
QY 2601 CTTATCATCAGAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCAAACATACCCG 2660  
Db 563 IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLysLys 582  
QY 2661 TGGGCC-----CGGAGCCATGAAGACAGAGCGAA---AGGACGACTGG 2702  
Db 583 TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602  
QY 2703 CAGAGCGCGCACACTGAGATCGTCGCCATCGACGCCCTCCACTTCAGAGCG----- 2753  
Db 603 GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622  
QY 2754 -----TACCTCGACAGTTTGTGCCGAGAGATGACAGCGGAGCTTAACAGGCTTAC 2807  
Db 623 AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysAlaSer 642  
QY 2808 TGTGATTTCTCGTCCCTGGAGTTTCTTCAGAGAACCTGTCTGACGTGCTACAGAAAC 2867  
Db 643 IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTyr 661  
QY 2868 TGGGCTGTGGTCTTTTGGGGGTGATGCTAGACTAAAAGCCTTAATACAGATCTCGCA 2927  
Db 662 TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681  
QY 2928 GCTGCTAGTGTAGCAGAGACGTGTTTATTTCACCTTTGGGAGACTCAAGACTGATGAGA 2987  
Db 682 AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAla 701  
QY 2988 GACATTTACAGATGCATACATCTCTCACTGAGAGAAACTGACTGTTGGAGAA 3041  
Db 702 LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719





Thu May 27 09:55:51 2004

Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aecmica-X-1  
CURRENT APPLICATION NUMBER: US 09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34248  
LENGTH: 2665  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL034555.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 9.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12  
OTHER INFORMATION: EST HUMAN HIT: AUL17052.1, EVALUATE 0.00e+00  
OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUATE 3.00e-10  
US-09-864-761-34248  
Alignment Scores:  
Pred. No.: 0.000342 Length: 2665  
Score: 184.00 Match: 126  
Percent Similarity: 33.69% Conservative: 92

Best Local Similarity: 19.47% Mismatches: 268  
Query Match: 2.54% Indels: 161  
DB: 9 Gaps: 24  
US-09-302-812-1 (1-4070) x US-09-864-761-34248 (1-2665)  
QY 6 AAAGTGAACGAGCCGGAATCAGAACGCTCATCTGAGGCTGGT----- 50  
DB 1381 ArgThrArgArgAsnValArgSerValTyrAlaThrMetGlyAspHisGluAsnArg 1300  
QY 51 -----AGGTGTCGGTGAAGAGGAGGAGCGGCTGGATAGGGCTGCTCGG 101  
DB 1301 SerProValLysGluProValGluGlnProArg-----ValThrArg 1314  
QY 102 GAGGTGTCAGAGCAGGAGCTGCAGAACGACATCAGCGGAGGAGGCGCATGTGTCGGGA 161  
DB 1315 LysArgLeu-GluArgGluLeuGlnGluAlaAlaValPro-----Th 1329  
QY 162 GGCACCGAGGAGGCGCGCTCCCTCCCTCCAGGGTGTG-----AA 206  
DB 1329 rThrProArgArgGlyArgProProLysThrArgArgAlaAspGluGluGluAs 1349  
QY 207 TGAGGCTCTACGCCG-----GGCTGCGCGCGGAGACTCACT 242  
DB 1349 nGluAlaLysGluProAlaGluThrLysProProGluGlyTrp--ArgSerProArg 1368  
QY 243 GCTGGGCTCCAGCATGAGTGGCGGCGCGCTGTGAGCCCTGCACCAAGCGACCCCGC 302  
DB 1369 SerGlnLysThrAlaAlaGlyGlyGlyProGlnGlyLysLysGlyLysAsnGluProLys 1388  
QY 303 TGGGACGCGCTGCACTTCTCCGCGCGCGCTCGGAGCGCGGAGCTTCCCGGCGAG 362  
DB 1389 ValAsp-----AlaThrArgProGluAlaThrThrGluVal-----GlyPro 1402  
QY 363 CAGAGCGCGCTCTCGATTCCCAAGAGCGCTCCGCTGAGTTCAGGCTCCCGCTCTCG 422  
DB 1403 GlnIleGlyValLysGluSerSerMetGluPro-----LysAlaAlaGluGluGlu 1419  
QY 423 TCAGGCTGGCGCTGGCGCGCGGAGCAGCACCGAGCGCGCGCCACTCTCTGTTTTC 482  
DB 1420 AlaGlySerGluGlnLysArgAspArgLysAspAlaGly----- 1432  
QY 483 AAACAGAGACTATACCACTGGATGGATGACACTAAAGGAATCAAGACAGTTGAATCAGAA 542  
DB 1433 -----ThrAspLysAsnProProGluThrAlaProValGluValValGluLysLys 1449  
QY 543 AGTTTCATAGTAAGAAACAAACAAATACAGAGAGAAATCCATGATGAGTCTGTACAA 602  
DB 1450 ProAlaProGluLysAsnSerLysSerLysArgGlyArgSerArgAsnSerArgLeuAla 1469  
QY 603 AAAGATAACTTTTATCAACATAACATGGAATAATAGAAATCTTTCTAGCTAGGTTT 662  
DB 1470 ValAsp-----LysSerAlaSerLeuLysAsnValAspAlaAlaValSer 1484  
QY 663 GATAAGTCACCATTGAAAAGGTACACAGTATTGAAAGCAGCATCAGACTGGGCTATG 722  
DB 1485 ProArgGlyAlaAlaAlaGlnAlaGlyGlu-----ArgLysSerGlyValValAlaVal 1502  
QY 723 TGTAAGTGGCAGAAATGAAGGCGCACACTCAGAA-----CGGCTTTTGGAAAGT 770  
DB 1503 SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer 1522  
QY 771 GAACCTCCAGCGTAACCTGTTACAGAGCAGTTCAGTAATGCTATGTCGATCAGTCG 830  
DB 1523 AspProValAspProAspLysGluProGluLys-----GluAspValSerAlaSer 1539  
QY 831 TCCCAAGAGTATCATCAGCTGACACAAATAGTGAGGAGAGTAGAGAGATAATACAGCTTT 890  
DB 1540 GlyProSerProGlu-----AlaThrGlnLeuAlaLysGlnMetGluLeuGluAla 1557  
QY 891 TTGACACATGTA---AAGCTTGCAGATGCAAGCAGACGATGGAAGATGAACAGGCGAGA 947  
DB 1558 ValGluHisIleAlaLysLeuAlaGluAlaSer----- 1568







QY	102	GAGCGTGTCCAGACGAGCAGCTGCAGAAGCAGTCAGCGGCAGAGGGGCCATGTCGCGGGA	161
Db	1932	LysArgLeu-GluArgGluLeuGlnGluAlaAlaValPro-----Th	1946
QY	162	GGCACCGAGAGGGGGCGGAGTCCGTCCTCCCAGGGTTAGTG-----AA	206
Db	1946	rThrProArgGlyVarGProProllysThrArgArgAlaAspGluGluGluAs	1966
QY	207	TGAGGCTCTACGCCCG-----GGTCGCCCGGAGACTCACT	242
Db	1966	nGluAlaLysGluProAlaGluThrLeuLysProProGluGlyTrp-ArgSerProArg	1985
QY	243	GCTCGCGGTCCCAGCATGAGTGGCGGCCCGGCTGTGAGCCTGCACCAGGACCCCCG	302
Db	1986	SerGlnLysThrAlaAlagIyGlyPrGclnglLyLysGlyAsnGluProllys	2005
QY	303	TGGAGCGCCTCGAATTCTCCGCGCGCCTCGACGCCCGGAGCTTCCCGCGCAGG	362
Db	2006	ValASP-----AlaThrArgProGluAlaThrThrGluVal-----GlyPro	2019
QY	363	CAGAGCGCGTCTCGATTCCAAGAGCGTCCGGTCGACTCAGGTCGCCGCGTCTCG	422
Db	2020	GlnIleelyValLysGluSerSerMetGluPro-----LysAlaAlaGluGluGlu	2036
QY	423	TCAGGTCGCGCCTCGGCGCGGCGGACACGACCGAGGACGCCACCTCTCTTGTTC	482
Db	2037	AlaGlySerGluGlnLysArgAspArgLysAspAlacly-----	2049
QY	483	AAACAGAAGACTATAACCACTGGATGGACACTAAAGGAATCAAGACAGTTGAAT	542
Db	2050	-----ThrAspLysAsnProProGluThrAlaProvalGluValValGluLysLys	2066
QY	543	AGTTTTGCATAGTAAGAAGAAAACAACATACAGAAGAATPCCATGATGATCTGTACAA	602
Db	2067	ProAlaProGluLysAsenSerLysArgLysArgSerArgAsnSerArgLeuAla	2086
QY	603	AAGATAACTTTTATCAACATAACATGGAAAAATTAGAAAATTGTCTCAGCTAGTTT	662
Db	2087	ValASP-----LysSerAlaSerLeuLysAsnValAspAlaAlaValSer	2101
QY	663	GATAAGTCAACAGTTGAAAAGGTACACAGTATTGGAAGCAGCATCAGACTGGGTATG	722
Db	2102	ProArgGlyAlaAlaAlaGlnAlaGlyglu-----ArgGluSerGlyValValAlaVal	2119
QY	723	TGTAAGTGGCAATGAAGGCCACACTCAGAA-----CGGCTTTTGGAAAGT	770
Db	2120	SerProGluLysSerGluSerProGlnLysGluAspGlyLeuSerSerGlnLeuLysSer	2139
QY	771	GAACCTCCAGGGTAACTCTGGTACACAGAGCTTCAGTAATGCTAATGCTCAGTCG	830
Db	2140	AspProValAspProAspLysGluProGluLys-----GluaspValSerAlaser	2156
QY	831	TCCCAAAGGATGATCAGGTGACACAATAAGTAGTGAGGAGTAGTAGATAATCAGCAGTTT	890
Db	2157	GlyProSerProGlu-----AlaThrGlnLeuAlalysGlnMetGluLeuGlnAla	2174
QY	891	TTGCACATGTA---AAGCTTGGCAATGCCAAGCAGACGATGGAAGATGAACAGGGCAGA	947
Db	2175	ValGluHisIleAlaLysLeuAlaGluAlaser-----	2185
QY	948	GAAGCCAGAACCCAGAGTGGCGAGCTTGCCATCCTCGAGAAGCCTGTGCGAGG	1007
Db	2186	-----AlaserAlalaTyrlsAlasAlaProGlugly	2197
QY	1008	TGTCAGCAGGAGGACAGACGTGTGTGCCGAGCCCTTGTCCGACACTGGCTCTGAG	1067
Db	2198	LeuAlaProGluuAsparGasp-----LysProAlaHisGlnAlaSerGluThr	2213
QY	1068	GATGTGTGTACTGCAGCTGAAAAT---GCCAACAGATTGAATACAGAAAGTAGTCTA	1124
Db	2214	GluLeuAlaAlaalaledlySeriIeileAsnAspileSerGlyGluProGluAsnPhe	2233
QY	1125	GGAAATTCCTCCATTGGAGAAAGTAGTAA-----	1157

Db	2234	ProAlaProProProfyzProGlyGluSerGlnThrAspLeuGlnProProAlaGlyAla	2255
QY	1158	-----CCTGAGTCACCAATGGATGTAGATAATTCCAAAANTAGTTGTCAG	1202
Db	2254	GlnAlaLeuGlnProSerGluGluGlyMetGluThrAspGluAlaValSerGlyIleLeu	2273
QY	1203	GATTCAGACGACATGAAGACACAGTCAGTTCAGGTTTGTATGACACAGAGAATACGACTTCT	1262
Db	2274	GluThrGluAlaThrGluSerSerArgProProValAsnAlaProAspProSerAla	2293
QY	1263	GCTCAACACGAAATAAACCTTCAAGGTTCCAAACACAGAGAGCTGACACTGAGTTGAGG	1322
Db	2294	GlyProThrAspThrLysGluAlaArgGlyAsnSerSerGluThrSerHisSerValPro	2313
QY	1323	AAGCGTCTCTGCTAAGGAGGTGAGATTCGATTACATTTCCAAATTGAGGA	1376
Db	2314	GluAlaLysGlySerLysGluValGluValThrLeuValArgLysAspLysGlyArgGln	2333
QY	1376	-----	1376
Db	2334	LysThrThrArgSerArgArgLysArgAsnThrAsnLysValValAlaProValGlu	2353
QY	1377	-----GGAGAGAGTCGAGCTGCAATGAATGAATGATGTG	1406
Db	2354	SerHisValProGluSerAsnGlnAlaGlnGlyGluSerProAlaAlaAsnGluGlyThr	2373
QY	1407	AATGCCAAACGACTGGAGTACTTCTAGCCTGAATGTAGAGTGCAGAAATTTCTAAGCAA	1466
Db	2374	ThrValGlnHisProGluAlaProGlnGlu-----GluLysGlnSerGluLysPro	2390
QY	1467	CATGGGAGAAAGGATCTTAAATCACAGATCATTTTCATGAGAGTGCCTCAAGAGAGAG	1526
Db	2391	HisSerThrProProGlnSerCysThrSerAspLeuSerLysIleProSerThrGluAsn	2410
QY	1527	AAAAAGAAAGAACAAATGTCAAATGAAACATCAAGAGAACAGAAAGAGATCCCTTAATAC	1586
Db	2411	SerSerGlnGluIleSerValGluGluArgThrProThrLysAlaSerValProProAsp	2430
QY	1587	ATTCCACCTCACCTTCTTCAGATAAGAAATGCTTGGAACTCCTATTGAGGAGATGAGG	1646
Db	2431	LeuProProProGlnPro-----AlaProValAspGlu-----	2442
QY	1647	AGATGCCAAGGTGTGGATCCGGCTG-----CCT	1676
Db	2443	---GluProGlnAlaArgPheArgValHisSerIleIleGluSerAspProValThrPro	2461
QY	1677	CCCTTGAGACCATCT	1691
Db	2462	ProSerAspProSer	2466

RESULT 12  
US-10-221-278-197  
; Sequence 197, Application US/10221278  
; Publication No. US20040034208A1  
; GENERAL INFORMATION:  
; APPLICANT: Hysseq, Inc  
; TITLE OF INVENTION: No. US20040034208A1el Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-045  
; CURRENT APPLICATION NUMBER: US/10/221,278  
; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: 09/693,267  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 09/665,363  
; PRIOR FILING DATE: 2000-09-19  
; PRIOR APPLICATION NUMBER: 09/616,847  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 09/596,193  
; PRIOR FILING DATE: 2000-06-17  
; PRIOR APPLICATION NUMBER: 09/574,454  
; PRIOR FILING DATE: 2000-05-19  
; PRIOR APPLICATION NUMBER: 09/519,705  
; PRIOR FILING DATE: 2000-03-07

; NUMBER OF SEQ ID NOS: 752  
; SEQ ID NO 197  
; LENGTH: 1163  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-221-278-197

Alignment Scores:  
Pred. No.: 0.00105 Length: 1163  
Score: 175.50 Matches: 152  
Percent Similarity: 33.8% Conservative: 108  
Best Local Similarity: 19.7% Mismatches: 256  
Query Match: 2.42% Indels: 252  
DB: 12 Gaps: 34

US-09-302-812-1 (1-4070) x US-10-221-278-197 (1-1163)

QY 59 GGTGAAGAGGAGGAGCGCTGG-----ATAGGCGCTGGTTCCGGAGGCTGTCTAG 112  
DB 100 GlyGlySerHisGlnSerLysThrProValGlyProAlaProSerThrSerGln 119  
QY 113 AGCAGGAGCTGCAGAGCAGTCCGCGCAGAGGGGCGCATGGTCCGGGAGGACCGCAGGA 172  
DB 120 SerGlnLys---ArgSerSerGlyLeuGlnSerGlyHisSerSerGlnArg-ThrSerAl 138  
QY 173 GGGGGC----- 179  
DB 138 aGlySerSerSerGlyThrAsnSerSerGlyGlnArgHisAspArgGluSerTyrAsnAs 158  
QY 180 -----CAGTCCGCTCCCTCCAGGCTTAGTGAATGAGG 211  
DB 158 nSerGlySerSerArgLysLysGlyGlnHisGlySerGluHisSerLysSerArgSe 178  
QY 212 CTCTAGCCCGGCTGGCCGAGACTCAGTGTGGGCTCCAGCATGAGTCGGGCCCC 271  
DB 178 rSerSerProGlyLysProGlnAlaValSerSerLeuAsnSerSerHisSerArgSerHi 198  
QY 272 CGGCTGTGAGCCCTGCACAGGACCCCGCTGGGACGGCGTGCACACTCTCCGCGCGC 331  
DB 198 sGlyAsnAspHisHisSerLysGluHisGlnArgSerLysSerProArgAspProAspAl 218  
QY 332 CGCCTCGGAGCCCGGAGCTTCCCGGAGGAGGCGGCTCTCGATCCAAAGGACGC 391  
DB 218 asnTrpAspSerProSer----- 224  
QY 392 TCGGTGCAGTTCAGGTCCTCCGCTCGTCCAGCTCGCCCTGGCGGGCGGGGACA 451  
DB 225 -----ArgValProPheSerSer-----GlyGl 232  
QY 452 GCAC-----CGAGGAGCGCCACCTCTCTGTTTTCAA 484  
DB 232 nhisSerThrGlnSerPheProProSerLeuMetSerLysSerAsnSerMetLeuGlnLy 252  
QY 485 ACAGAGACTATAACACGAGTGGATGACACTAAGGAATCAAGACAGTTGAA----- 536  
DB 252 sProThrAlaTyrValArgProMetAsp-----GlyGlnGluSerMetGluProLysLe 270  
QY 537 -----TCAGAAAGTTTCATAGTAAGAAACAACATACAGAGAGAAATCCATGATGAG 592  
DB 270 uSerSerGluHisTyrSerSerGlnSerHisGlyAsnSerMetThrGluLeuLysProSe 290  
QY 593 TTCTGTACAAAAGATAACTTTTATCAACATACATGGAATAATTAGAAATGTTTCTCA 652  
DB 290 rSer-----LysAlaHisLeuThrLysLeuLysIleProSerGI 303  
QY 653 G-----CTAGGTTTGTAGTACCGAGTTGAAAGAGGTACACGATTTT 697  
DB 303 nProLeuAspAlaSerAlaSerGlyAspValSerCysValAsp-----GluIleLe 320  
QY 698 GAAGCAGCATCAGATCGCGCTATGTGTAAAGTGGCAGATGAAGGGCCACATCAGAACG 757  
DB 320 uLysGlu----- 322

QY 758 GCTTTTGGAAAGTCAACCTCCAGCGGTAACTCTCGTA----- 794  
DB 323 -MetThrHisSerTrpProProLeuThrAlaIleHisThrProCysLysThrGluPr 342  
QY 794 ----- 794  
DB 342 oSerLysPheProPheProThrLysGluSerGlnGlnSerAsnPheGlyThrGlyGluGl 362  
QY 795 -----CCAGAGCAGTTTCAGTAATGCTAATGCTCGATCAGTCCGCCCAAGGA 841  
DB 362 nLysArgTyrAsnProSerLysThrSerAsnGlyHisGlnSerLysSerMetLeuLysAs 382  
QY 842 TGATCAGAGTGCACAAATAGTAGGAGAGTAGAGATAATCAGCAGTTT---TTGACACA 898  
DB 382 pAspLeuLysLeuSerSerGluAspSerAspGlyGluGlnAspCysAspLysThrMe 402  
QY 899 TGTAAGCTTGGAAATGCAAGCAGACGATGGAAGATGAACAGGCGCAGAGCCAGAG 958  
DB 402 tProArgSerThrProGlySerAsnSerGluProSerHisHisAsnSerGluLysAlaAs 422  
QY 959 CCACAGAGTGTGGCAGGCTTGCATCCT---GCAGAAGCCTCTGCAGGCGGTGTGACGA 1015  
DB 422 pAsnSerArgAspAspSerSerSerHisSerGlySerGluSerSerGlySerAspSe 442  
QY 1016 GGAGGAGACAGACGCTGGTCCGAGAGCCCTTGTTCGACACTGGCTCTGAGGATTTGG 1075  
DB 442 rGlu-----SerGluSerSerSerAspSerGluAlaAsnGluProSe 457  
QY 1076 TACTGGA-----CTGAAAAT--- 1091  
DB 457 rGlnSerAlaSerProGluProGluProProThrAsnLysTrpGlnLeuAspAsnTr 477  
QY 1092 -GCCAACAGATTGAATAGACAA-----GAAAGTAGTCTAGG 1126  
DB 477 pleuAsnLysValAsnProHisLysValSerProAlaSerSerValAspSerAsnIlePr 497  
QY 1127 AAATCTCTCCTCCTTGGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATATTC 1186  
DB 497 oSerSerGlnGlyTyrLysLysGlyLysGluGln-----GlyTh 511  
QY 1187 CAATAATAGTCTCAGGATTCAGACAGCAGATGAAGACAGCAAGT-----CCAGGTTT 1237  
DB 511 rGlyAsnSerTyrThrAspThrSerGlyProLysGluThrSerSerAlaThrProGlyAr 531  
QY 1238 TCATGAA-----CAGGAAGATAGCAGTCTCTCAACAGCAATAAACCCTTCAAG 1288  
DB 531 gAspSerLysThrIleGlnLysGlySerGluSerGlyArgGlyArgGlnLys----- 548  
QY 1289 GTTCCAAACCAAGAGAGCTGCACACTGAGTTGAGGAGCGGCTCTCTGCTAAGGAGGTGA 1348  
DB 549 ---SerProAlaGlnSerAspSerThrThrGlnArgArgThrValGlyLysGlnPr 567  
QY 1349 GATTCCATTACATTTCCCAATTTGAAGGAGGAGAGTCCGAGTGGAAATGAATGATGTGA 1408  
DB 567 oLysLysAlaGlu---LysAlaAlaAlaGluGluProArgGlyLysLys---IleGl 585  
QY 1409 TGCCAAACGACCTGGAAGTACTTCTAGCCTGAAATGATGAGTGCAGAAATTTCT----- 1460  
DB 585 uSerGluThrProValAspLeuAlaSerSerMetProSerSerArgHisLysAlaAlaTh 605  
QY 1461 -----AAGCAACATGGGAGAAAGATTCTAAATCAACAGATCAITTCATGAG 1507  
DB 605 rLysGlySerArgLysProAsnIleLysLysGluSerLysSerSerProArg----- 622  
QY 1508 AGTGCCCAAGCAGGACAAAGAAAGCAATGTGAATGAACATCAAGAACAGACA 1567  
DB 623 ---ProThrAlaGluLysLysLysLysLysLysSerLysSerLysSerLysSerAr 641  
QY 1568 AAGGAAGATC----- 1577  
DB 641 gGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerGluSerLeuProPr 661  
QY 1578 -----CCTAAATACATCCACCTCCTTCTCCAGATAAG----- 1613



1016	QY	GGAGGAGACAGCGTGGTGTCCGAGAGCCCTTGTGCGACACTCGCTCTCAGGATGTGG	1075
442	Db	 rGlu-----SerGluSerSerSerAspSerGluAlaAsnGluProSe	457
1076	QY	TACTGGA-----CTGAAAAAT--	1091
457	Db	rGlnSerAlaSerProGluProProProThrAsnLysTrpGlnLeuAspAsnTr	477
1092	QY	-----GCCAACAGATTGAATAGACAA-----GAAACTAGCTTAGG	1126
477	Db	pLeuAsnLysValAsnProHisLysValSerProAlaSerSerValAspSerAsnLlePr	497
1127	QY	AAATTCTCTCCATTGTGAAAGAAAGTGAACCTGAGTCACCAATGGATGTAGATAATTC	1186
497	Db	oSerSerGlnGlyTrpLysLysGluGlyArgGluGln-----GlyTh	511
1187	QY	CAAAAAATAGTTGTCCAGATTCCAGACGACGATGAGAGCAAGT-----CCAGGTTT	1237
511	Db	xGlyAsnSerTyThrAspThrSerGlyProLysGluThrSerSerAlaThrProGlyAr	531
1238	QY	-----TGATGAA-----CAGAAAGATGACAGTCTGCTCAACAGACAAATAAACCTTCAAG	1288
531	Db	gAspSerLysThrIleGlnLysGlySerGluSerGlyArgGlnLys-----	548
1289	QY	GTTCCAACCAAGAGAAAGCTCACACTGATGTGAGAGACCGTCTCTGTAAAGGAGGTGA	1348
549	Db	-----SerProAlaGlnSerAspSerThrThrGlnArgArgThrValGlyLysLysGlnPr	567
1349	QY	GATTCGATTATCATTTCCAAATTTGAAGAGAGAGAGTCCGAGCTGGATGAATGATGTGAA	1408
567	Db	oLysLysAlaGlu-----LysAlaAlaLagLiuProArgGlyGlyLeuLys---IleG	585
1409	QY	TGCCAAACGACCTGGAAGTACTCTACCTCGAATGTAGAGTCGCAAAATCTCT-----	1460
585	Db	uSerGluThrProValAspLeuAlaSerSerMetProSerSerArgHisLysAlaAlaTh	605
1461	QY	-----AAGCAACATGGGAGAAAGGATTTAAATACAGATCATTTTCATGAG	1507
605	Db	rLysGlySerArgLysProAsnLleLysLysGluSerLysSerSerProArg-----	622
1508	QY	AGTGCCCAAGAGGAGCAAAAGAAAGACATGTGAATGAACATCAAGAACAGACA	1567
623	Db	ProThrAlaGluLysLysLysTyLysSerThrSerLysSerSerGlnLysSerAr	641
1568	QY	AAGGAAGATC-----	1577
641	Db	gGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerLeuProPr	661
1578	QY	-----CCTAAATACATCCACCTCACCTTCTCCAGATAG-----	1613
661	Db	oSerSerGlnThrProLysTyLysProGluSerAsnArgThrProValLysProSerVa	681
1614	QY	-----AAATGGCTTGGAACTCCTATTGAGGAGATGAGGAG	1648
681	Db	lGluGluGluAspSerPhePheArgGluArgMetPheSerProMetGluLysGluLe	701
1649	QY	AATGCCAAGGTGTGGGATCCGGCTCGCTCCCTTTGAGACACCTGCCAATCACACGTGAC	1708
701	Db	uLeuSer-----ProLeuSerGluProAspAspArgTyrProLeuIl	715
1709	QY	TATTCGGGTAGATCTTTTGGCAATAGGAGAGTCTCT-----AAACCTTTC	1754
715	Db	eValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLysProTyrLysGluThrGl	735
1755	QY	-----CCACACACATTTTAAAGATTGTGGGACACCAAGCATGTT-----	1793
735	Db	uProProGlyGlyGluLysLysAsnValProGluLysHisThrArgGluAlaGlnLysGl	755
1794	QY	-----AAGATGCTTTGTTTCAGAACAAACATTTGCTACCTGTGGAGATGAGATGG	1843
755	Db	nAlaSerGluLysValSerAsnLysGlyLysArgLysHisLysAsnGluAspAspAsn--	774
1844	QY	TGACGGAGCTGCAGGACGCCGG	1865

```

Db      775 ----ArgAlaserGluserLys 780          |||||::: ||||:::
RESULT 14
US-10-287-218-17
; Sequence 17, Application US/10287218
; Publication No. US20030198975A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: AZIMZAI, Yalda; AU-YOUNG, Janice K.
; APPLICANT: BATRA, Sajeev; BAUGHN, Mariiah R.
; APPLICANT: BECHA, Shanya D.; BOROWSKI, Mark L.
; APPLICANT: BUFORD, Neil; DING, Li
; APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.
; APPLICANT: GANDHI, Ameena R.; GIETZEN, Kimberly J.
; APPLICANT: GRIFFIN, Jennifer A.; HAFALIA, April J.A.
; APPLICANT: HONCHELL, Cynthia D.; LAL, Preeti G.
; APPLICANT: LEE, Soo Yeun; LU, Dyung Aina M.
; APPLICANT: ARVIZU, Chandra S.; RAMKUMAR, Jayalaxmi
; APPLICANT: REDDY, Roopa; SANJANWALA, Madhu, M.
; APPLICANT: TANG, Yi Tom; WALIA, Narinder K.
; APPLICANT: WANG, Yu-mei, E.; WARREN, Bridget A.
; APPLICANT: XU, Yuming; YANG, Junming
; APPLICANT: YAO, Monique G.; YUE, Henry
; APPLICANT: ZEBARJADIAN, Yeganeh
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH
; FILE REFERENCE: PI-0417 USA
; CURRENT APPLICATION NUMBER: US/10/287,218
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: PCT/US02/11152
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: US 60/349,705
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/295,263
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/295,340
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/293,727
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/291,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/291,662
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: US 60/287,228
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/286,820
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/283,294
; PRIOR FILING DATE: 2001-04-11
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 17
; LENGTH: 1462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030198975A1 5734806CDI
US-10-287-218-17

Alignment Scores:
Pred. No.: 0.00135 Length: 1462
Score: 174.50 Matches: 158
Percent Similarity: 33.84% Conservative: 110
Best Local Similarity: 19.95% Mismatches: 298
Query Match: 2.41% Indels: 226
DB: 14 Gaps: 32

US-09-302-812-1 (1-4070) x US-10-287-218-17 (1-1462)
QY 113 AGCAGGAGCTGCAGAGCAAGTCAGTCCGCCAGGGCGCATGTGCGGGAGGCACCGAGGA 172
    ::: |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| |

```

511 Db ThrHisSerSerArgAspSerTyrArgSerLysSerHisSerGlnSerTyrSerArgGly 530  
 173 QY GGGGGCCAGTCCGTCCTCCAGGGTTACTGAATGAGGCTCTACAGCCCGGGCTGGCCCG 232  
 531 Db -SerArgSer-----ArgThrAlaSerLysSerSerHisSerArgSerAr 547  
 233 QY GAGACTCAGTGTCTGGGTCCAGCATGAGTGGCGGCCCGCGTGTGAGCCCTGCACCAA 292  
 547 Db gSerLysSerArg-----SerSerSerLysSerGlyHisArgLysArgAlaSerLy 564  
 293 QY GCGACCCCGTGGGACGCGCT-----CCAACTCTCCGCC 328  
 564 Db sSerProArgLysThrAlaSerGlnLeuSerGluAsnLysProValLysThrGluProLe 584  
 329 QY GCGCGCTCCGACCGCCGCGAC-----TTCCCGCGCAGCGACAG 367  
 584 Db uArgAlaThrMetAlaGlnAsnGluAsnValValGlnProValValAlaGluAsnIl 604  
 368 QY GCGCGTCTCCGATCCCAAGGACGCTCCGGTCCGTCAGTTCAGGCTCCGCGCTCCGTCAGG 427  
 604 Db eProValIleProLeuSerSerProProProSerArgTPrLysProGlyGlnLysPr 624  
 428 QY CTGC-----GCCCTGGCGCGGGACACACACGAGCGGACGCCACCTCTCTGTTT 481  
 624 Db oTrpLysProSerTyrGluArgIleGlnGluMetLysAlaLysThrThrHisLeuLeuPr 644  
 482 QY CAACAGAGACTATAACCTAGTTCGATGGACACTAAAGGAATCAAGCAGTTCGAATCAGA 541  
 644 Db oIleGlnSerThrTyr-----SerLeuAlaAsnLys-----GluThrGlySerSerSe 661  
 542 QY AAGTTTGCATAGTAAAGAAACAAACATACAGAGAGAAATCCATGATGATTCGTGA-- 599  
 661 Db rSerTyrHisLysArgGluLysAsnSerGluSerAspGlnSerThrTyrSerLysTyrSe 681  
 600 QY -----CAAAAGATACTTTA 616  
 681 Db rAspArgSerSerGluSerSerProArgSerArgSerArgSerArgSerArgSerTyr 701  
 617 QY TCAACATAACATGGAATAATAGAAATGTT--TCTCAGCTAGGTTTGTGATAAGTCACC 673  
 701 Db rSerArgSerTyrThrArgSerArgSerLeuAlaSerSerHisSerArgSerArgSerPr 721  
 674 QY AGTTGAAAAGGT-----ACACGATTTGTAAGCAGCATCAGACTCGCGCTATGTG 724  
 721 Db oSerSerArgSerHisSerArgAsnLysTyrSerAspHisSerGlnCysSerArgSerSe 741  
 725 QY TAAAGTGG-----CAGAATGAAGGCCACACTCAGAACGCTTTTGGAAAGTGA 772  
 741 Db rSerTyrThrSerIleSerSerAspAspGlyArgArgAlaLysArgArgLeuArgSerSe 761  
 773 QY ACCTCCAGCGGTAACTCTGTCACGAGCAGTTCAGTAAATGCTAATGTC----- 821  
 761 Db rGlyLysLysAsnSerValSerHisLysLysHisSerSerSerSerSerSerSerSerSer 781  
 822 QY -----GATCAGTGTGCCCCAAAGAGATGATCAGATGACACACAAA 859  
 781 Db sSerLysTyrValLysGlyArgAspArgSerSerCysValArgLysTyrSerGluSerAr 801  
 860 QY TAGT-----GAGGAGGTAGATATATCAGCAGTTTTGTACATGTAAGCTTCGAA 913  
 801 Db gSerSerLeuAspTyrSerSerAspSerGluGln-----SerSerValGlnAlaThrGl 819  
 914 QY TGCAGAGCAGCATGGAATGAACAGGCGCAGAGAGCCAGGACCCAGAGAGTGTGG 973  
 819 Db nSerAlaGlnGluLysGlu--LysGlnGlyGlnMetGluArgThrHisAsnLysGlnGl 838  
 974 QY CAGAGCTTGCATCTGCAGAACGCTGTGCAGGGTGTGCAGGAGGACACACCTGTGT 1033  
 838 Db uLys-----AsnArgGlyGluGluLysSerSerSerSer 849  
 1034 QY GTCCGAGAGCCCTTGTGCGGAC-----ACTGGCTCTGAGGATGTTGTGCTGAGCTGAA 1087  
 849 Db uArgGluCysProHisSerLysLysArgThrThrLeuLysGluAsnLeuSerAspHisLeuAr 869

1088 QY AATGCCCAACAGATTGAATAGACAAAGAGTAGCTAGGAAATTCCTCCATTTGAGAA 1147  
 869 Db gAsnGlySerLysProLysArgLys-----AsnTyrAlaGlySerLysTr 884  
 1148 QY AGAAAGTGAACCTGAGTCCACCAATGATGTA--GATAATTCAAAATAATAGTTGTACAGA 1204  
 884 Db pAspSerGluSerAsnSerGluArgAspValThrLysAsnSerLysAsnAspSerHisPr 904  
 1205 QY TTCAGAGCAGATGCAAGAGACAAAGTCCAGGTTTGTATGAACACAGAGATAGCAGTTCTGC 1264  
 904 Db oSerSerAspLysGluGluGluLysThrSerAspSerGluSerGluValSerGluIl 924  
 1265 QY TCAACAGCAAAATAAACCTTCAAGGTTCCCAACCAAGAGAGCTCACACTGAGTTGAGAA 1324  
 924 Db eHisIleLysValLysProThr-----Th 932  
 1325 QY GCGTCTCTGCTAAGGAGGTGAGATTTCGATTACATTTCCAATTTGAAGAGAGAGAG 1384  
 932 Db rLysSerSer-----Th 936  
 1385 QY TCGAGCTGGGAATGAATGATGAAT-----GCCNAACGACCTGGAAGTAC 1429  
 936 Db rAsnThrSerLeuProAspAspAsnGlyAlaTrpLysSerSerLysGlnArgThrSerTh 956  
 1430 QY TTCTAGCTGAATGTAGAGTGCAGAAATTCCTAAG----- 1463  
 956 Db rSerAspSerGluGlySerCysSerAsnSerGluAsnAsnArgGlyLysProGlnLysHi 976  
 1464 QY CAACATGGGAGAAAGATTCATAAATCACAGATCATTTTCATGAGGTGCCCAAGCAGA 1522  
 976 Db sLysHisGlySerLysGluAsnLeuLysArgGluHis-----ThrLysLysValLy 993  
 1523 QY GCACAAAGAAAGAAACAATGTGAATCAACATCAAGACACAAAGAGAGATCCCTAA 1582  
 993 Db sCLeuLysGlyLysLysAspLysHisLysAlaProLysArgLysGlnAlaPh 1013  
 1583 QY ATACATTCACCTCACCTTCTCCAGATAGAAATGCTTCGAACCTCCTATTGAGGAGAT 1642  
 1013 Db eHisTrpGlnPro----- 1017  
 1643 QY CAGGAGAAATGCAAGGTGGGATCCCGCTCCCTTGGAGACCATCTGCCAATCACAC 1702  
 1017 Db ----- 1017  
 1703 QY AGTCACTATTCGGTAGATCTTTTCCGNATAGGAGAGCTTCTTAACCTTTCCCAACACA 1762  
 1018 Db -----ProLeuGluPheGlyGluGl 1024  
 1763 QY TTTTAAAGATTGTGGGCAACAACAGCATGTTAAGATGCTTTGTCAGAACAAACTGTGA 1822  
 1024 Db uGluGluGluLysLysAspLysGlnValThrGlnGluSerLysGluLysLysVal-- 1043  
 1823 QY CCCTGTGGAAGATGAGATGTGAGGAGCTGCAGGACCGGTGGGAACTCATTCAGAC 1882  
 1044 Db -----SerGluAsnAsnGlu----- 1048  
 1883 QY TGCATCTTCAACAGGCTCACTCGGCCCCAGAACCTGAGGATGCTATTCTCAAGTACAA 1942  
 1049 Db -----ThrIleLysAspAsnIleLeuLysThrGl 1058  
 1943 QY TGTGGCA-----TATTCTAAGAAATGGGACTTTACAGCTTTGATTGATTCTGTG 1990  
 1058 Db uLysSerSerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLue 1078  
 1991 QY GGATAAGGTACTAGAAAGAGAGAGAGCTCAACACTGTATCATCTTCCTCCCTGATAT 2050  
 1078 Db uAspGlnPheThrLysAspAspSerLysLeuSerLysProThrAlaLeuAsnThrGl 1098  
 2051 QY GGTGAAATTCAGCTCTGCTCCAAATATTGTTACCCAGCCAATACCACTCCTCGAACA 2110  
 1098 Db uGluAsnValAlaA---CysLeuGlnAsnIle-----Gl 1108

QY 2111 GAAGATGAATCATTCATCAATGTCACAGAACAGATTCGCCAGTCTTTTAGCTTAATGC 2170  
 Db 1108 nHisValGluSerValProAsnGlyValGluAspValLeuGlnThrAspAspAsnMe 1128  
 QY 2171 TTTCTTCGACGTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGT----- 2225  
 Db 1128 tGluileCysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLe 1147  
 QY 2226 -----TATCCAGATATTAAAC 2240  
 Db 1147 uGlyAsnAlaArgLeuAspThrProAspIleAsn 1158

RESULT 15  
 US-09-932-257A-3  
 ; Sequence 3, Application US/09932257A  
 ; Publication No. US20030039658A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Roeder, Robert  
 ; APPLICANT: Estable, Mario  
 ; TITLE OF INVENTION: MCF7, A No. US20030039658A1el Transcription Factor  
 ; FILE REFERENCE: 600-1-269N  
 ; CURRENT APPLICATION NUMBER: US/09/932,257A  
 ; CURRENT FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/226,340  
 ; PRIOR FILING DATE: 2000-08-18  
 ; PRIOR APPLICATION NUMBER: 60/226,339  
 ; PRIOR FILING DATE: 2000-08-18  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 3  
 ; LENGTH: 1163  
 ; TYPE: PRN  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (1086)..(1086)  
 ; OTHER INFORMATION: x is any amino acid  
 US-09-932-257A-3

Alignment Scores:  
 Pred. No.: 0.00147 Length: 1163  
 Score: 173.50 Matches: 150  
 Percent Similarity: 31.91% Conservative: 96  
 Best Local Similarity: 19.46% Mismatches: 267  
 Query Match: 2.40% Indels: 258  
 DB: 10 Gaps: 33

US-09-302-812-1 (1-4070) x US-09-932-257A-3 (1-1163)

QY 59 GGTGAAGAGGAAGCAGCGCTCTCG-----ATAGGGCGCTGGTTCCGGAGGCTGTCTAG 112  
 Db 100 GlyGlySerHisGlnSerLysTrpThrProValGlyProAlaProSerThrSerGln 119  
 QY 113 AGCAGAGTGCAGAGAGCTAGCGGAGGAGGCGATGTCGCCGGGAGGACCGAGGA 172  
 Db 120 SerGlnLys---ArgSerGlyLeuGlnSerGlyHisSerGlnArg-ThrSerAl 138  
 QY 173 GCGGGCG----- 179  
 Db 138 aGlySerSerSerGlyThrAsnSerSerGlyGlnArgHisAspArgGluSerTyrAsnAs 158  
 QY 180 -----CAGTCGCTCCCTCCACAGGCTTAGTGAATGAGG 211  
 Db 158 nSerGlySerSerSerArgLysLysGlyGlnHisGlySerGluHisSerLysSerArgSe 178  
 QY 212 CTCTACGCCCGGCTGCCCGGAGACTCATGCTGCGGGTCCAGCATGATGTCGGGCCCC 271  
 Db 178 rSerSerProGlyLysProGlnAlaValSerSerLeuAsnSerSerHisSerArgSerHi 198  
 QY 272 CGGCTGTGACCCCTGCACCAAGGACCCCGCTGGGACCGCGCTGCAACTCTCCGCGGCG 331  
 Db 198 sGlyAsnAspHisHisSerLysGluHisGlnArgSerLysSerProArgAspProAspAl 218

QY 332 CGCTCTCGACGCCCGAGAGTTCCTCCCGGAGCGCAGAGCGCGGTCTCTCGATTCCAGGACGC 391  
 Db 218 aAsnTrpAspSerProSer----- 224  
 QY 392 TCCGGTGCAGTTTCCAGGTCCTCGCGTCTCTCTCGAGTCTCGCTCCGCGCGGCGGAC 451  
 Db 225 -----ArgValProPheSerSer-----GlyGl 232  
 QY 452 GCAC-----CGAGGCAGCGCCACCTCTCTTTGTTTCAA 484  
 Db 232 nHisSerThrGlnSerPheProSerLeuMetSerLysSerAsnSerMetLeuGlnLys 252  
 QY 485 ACAGAAGACTATACCACTGGATGGACACTAAAGGAATC-----AGACAGATTGAATC 538  
 Db 252 sProThrAlaTyrValArgProMetAspGlyGlnGlySerMetGluProLysLeuSerSe 272  
 QY 539 AGAAGTTTGCATAGTAAAGAAACAAACATACAGAGAGAAATCCATGATGAGTCTTGT 598  
 Db 272 rGluHisTyrSerSerGlnSerHisGlyAsnSerMetThrGluLeuLysProSerSer-- 291  
 QY 599 ACAAAAAGATAACTTTTATCAACATACATGGAATAATAGAAAATGTTTCTCAG----- 653  
 Db 292 -----LysAlaHisLeuThrLysLeuLysLysLysLysLysLysLysLysLysLys 305  
 QY 654 -----CTAGGTTTGTATAGTACACAGTTCAAAAGAGTACACAGTATTTCAGCA 703  
 Db 305 uAspAlaSerAlaSerGlyAspValSerCysValAsp-----GluileLysGl 322  
 QY 704 GCATCAGACTCGGCTATGTGTAGTGGCAGAAATGAAGGCGCACACTCAGAACCGGCTTTT 763  
 Db 322 u-----Meth 324  
 QY 764 GGAAAGTGAACCTCCAGCGGTAACTCTGTA-----CCAGACA 802  
 Db 324 rHisSerTrpProProLeuThrAlaIleHisThrProCysLysThrGluProSerLys 344  
 QY 803 GTTC-----AGTAATGCTAATGTCATGATCAGTCGTCCTCC 835  
 Db 344 sPheProPheProThrLysGluSerGlnGlnSerAsnPheGlyIleGlyGluGlnLysAr 364  
 QY 836 AAAGGATGATCAGTGCACAAATAGTGCAGAGAGTAGAGATAATACAGACAGTCTTTGAC 895  
 Db 364 gTyrAsnProSerLysThrSerAsnGlyHisGlnSerLys-----SerMetLeuLysAs 382  
 QY 896 ACATCTAAAGTTCGAAATGCAAGCAGACGATGGAAGATGAACAGGCGCAGAGACCCAG 955  
 Db 382 pGlyLeuLysLeuSerSerSer-----GluAspSerAspGly----- 394  
 QY 956 AGCCACAGAGTCTGGCAGGCTTGC----- 983  
 Db 395 -----GluGlnAspCysAspLysThrMetProArgSerThrProGlySerAsnSerGluPr 413  
 QY 984 -----CATCTCTCAGAAAGCCTGTGCAGGGGTGTGCAGAGGAGGACAGACGTTGTCT 1036  
 Db 413 oSerHisHisAsnSerGlu-----GlyAlaAspAsnSerArgAspAspSerSerSe 430  
 QY 1037 CGAGAGCCCTTGTCCGACATGGCTCTCAGGATGTTGTTACTGACTGAAATAATCCCAA 1096  
 Db 430 rHisSerGlySerGluSerSerSerGlySerSerGluSerGluSerSerSerSerSerSe 450  
 QY 1097 CAGATTGAATAGACAAAGAAAGTAGTCTAGGAAATCTCCTCCATTTTGAAGAAAGAGTGA 1156  
 Db 450 p-----SerGluAlaAsnGluProSerGlnSerAlaSerProGl 463  
 QY 1157 ACCTGAGTACCA-----ATGGATGTAGATAAT----- 1184  
 Db 463 uProGluProProProThrAsnLysTrpGlnLeuAspAsnTrpLeuAsnLysValAsnPr 483  
 QY 1184 ----- 1184  
 Db 483 oHisLysValSerProAlaSerSerValAspSerAsnIleProSerSerGlnGlyTyrLy 503  
 QY 1185 -----TCCAAAATAGTTCTCAGGATTTCAGAGCAGATGAAGA 1222



```
Db      503 sLysGluClyArgGluGlnGlyThrGlyAsnSerTyrThrAspThrSerGlyProLysGlu 523
Qy      1223 GACAAGT-----CCAGGTTTTGATGAA-----CAGGAAGATAGCAGTTCTGC 1264
Db      523 uThrSerAlaThrProGlyArgAspSerLysThrIleGlnLysGlySerGluSerGlu 543
Qy      1265 TCAACACAGCAATAAACCCTTCAAGGTTCCAACAGAGAGAGCTGACACTGAGTTGAGGAA 1324
Db      543 yArgGlyArgGlnLys-----SerProAlaGlnSerAspSerThrThrGlnAr 559
Qy      1325 GCGGTCTCTCTCTAAGGAGGTGAGATTTCGATTACATTTCCAAATTTGAAGAGAGAGAG 1384
Db      559 gArgThrValGlyLysLysGlnProLysLysAlaGlu---LysAlaAlaAlaGluGluPr 578
Qy      1385 TCGAGCTGGAATGATGATGATGATGATGCCAACGACCTGGAGTACTCTTACCTGCTGAATGT 1444
Db      578 oArgGlyGlyLeuLysIleGluSerGluThrProValAspLeuAlaSerSerMetProSe 598
Qy      1445 AGAGTGCAGAAATTCCTAAGCAACATGGGAGAAAGGATTCTAAATATCACAGATCATTTTCAT 1504
Db      598 rSerArgHisLysAlaAlaThrLysGlySerArgLysProAsnIleLysLysGluPheLy 618
Qy      1505 GAGAGTGCCAAA---GCAGAGGACAAAGAAAGAACAAATGTGAAATGAAACATCAAAG 1561
Db      618 sSerSerProArgProThrAlaGluLysLysLysTyrLysSerThrSerLysSerSerGlu 638
Qy      1562 AACAGAAAGGAGATC----- 1577
Db      638 nLysSerArgGluIleIleGluThrAspThrSerSerSerAspSerAspGluSerGluSe 658
Qy      1578 -----CCTAAATACATTCACCTCACCTTCTCCAGATAAG-- 1613
Db      658 rLeuProProSerSerGlnThrProLysTyrProGluSerAsnArgThrProValLysPr 678
Qy      1614 -----AAATGGCTTGGAACTCCTATTGAGGA 1639
Db      678 oSerSerValGluGluGluAspSerPheArgGlnArgMetPheSerProMetGluGl 698
Qy      1640 GATGAGGAGAAATGCCAAGGTGTGGGATCCGGCTGCCTCCCTTGAGACCATCTGCCAATCA 1699
Db      698 uLysGluLeuLeuSer-----ProLeuSerGluProAspAspArgTy 712
Qy      1700 CACAGTGACTATTCGGGTAGATCTTTTGCATATAGGAGAAGTTCTCT---AAACCTTTC-- 1754
Db      712 rProLeuIleValLysIleAspLeuAsnLeuLeuThrArgIleProGlyLysProTyrLy 732
Qy      1755 -----CCAACACATTTTAAGATTGTGGGACACAAAGCATGTT----- 1793
Db      732 sGluThrGluProProLysGlyGluLysLysAsnValProGluLysHisThrArgGluAl 752
Qy      1794 -----AAGATGCTTGTTCAGAACAAACAACTGTACCTGTGGAAGA 1834
Db      752 aGlnLysGlnAlaSerGluLysValSerAsnLysGlyLysArgLysHisLysAsnGluAs 772
Qy      1835 TGAGAATGTCAGCAGCTGCAGGCGCGG 1865
Db      772 pAspAen-----ArgAlaSerGluSerLys 780
```

Search completed: May 26, 2004, 18:34:40  
Job time : 252.229 secs



Db	1	MetSerAlaGlyProGlyCysGluProCysThrIysArgProArgTrpAspAlaAla	20
Qy	318	ACTTCTCCGCGCGCCCTCGAGCCGCGAGCTTCCCGCGAGCAGAGCGCGTCTC	377
Db	21	ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValIleu	40
Qy	378	GATTCCAAAGGAGCTCCCGGTGAGTTCAAGGTCCCGCGCTCTCTGTCAGGTGCGCCCTG	437
Db	41	AspSerIysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaIleu	60
Qy	438	GGCGGGGGGACAGCAGCCGAGGAGCCACCTCTCTGTTTCAACAGAACTATA	497
Db	61	GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle	80
Qy	498	ACCAGTTGGATGGACATAAAGGAATCAAGACAGTTGAATCAGAAAGTTTCATAGTAA	557
Db	81	ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerIys	100
Qy	558	GAACAAACAAATACAGAGAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTTAT	617
Db	101	GluAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr	120
Qy	618	CAACATAACATGAAAAATTAGAAATGTTTCTCAGCTAGGTTTTGATAAGTCAACAGTT	677
Db	121	GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal	140
Qy	678	GAATAAGGTACACAGTATTGAAGCAGCATCAGACTCGCGGTATGTGAAGTGGCAGAT	737
Db	141	GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn	160
Qy	738	GAAGGCCACACTCAGACCGCTTTTGGAAAGTGAACCTCCAGCGTAACTCTGTACCA	797
Db	161	GluGlyProHisSerGluArgLeuLeuGluSerGluProAlaValThrLeuValPro	180
Qy	798	GAGAGTTTCAGTAATGCTAATGCTGATCAGTCTGCTCCCAAGGATGATCAGACTCACACA	857
Db	181	GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr	200
Qy	858	AATAGTGAGGAGTAGAGATAATCAGCAGTTTTTGACACATGTAAAGCTGCGAATGCA	917
Db	201	AsnSerGluGluSerArgAspAsnGlnPheLeuThrHisValLysLeuAlaAsnAla	220
Qy	918	RAGCAGCATGAGATGACAGCGCCAGAGAGCCAGAGCCACAGAGTGTGGCAG	977
Db	221	LysGlnThrMetGluAspGlnGlnGlyArgGluAlaAlaArgSerHisGlnLysCysGlyLys	240
Qy	978	GCTTGCCATCTCGCAGAGCCTGTGCGGGTGTGAGGAGGAGAGCAGCGTGTGTCC	1037
Db	241	AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer	260
Qy	1038	GAGAGCCCTTGTGCGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAATGCCAAC	1097
Db	261	GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn	280
Qy	1098	AGATTGATAGACAGAAAGTAGTCTAGAAATCTCTCCATTGAGAGAAAGAGTGA	1157
Db	281	ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProPheGluLysGluSerGlu	300
Qy	1158	CTGTAGTCACCAATGGATGATGATTAATCCAAAAATAGTTGTGAGATTTCAGAACAGAT	1217
Db	301	ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp	320
Qy	1218	GAAGAGCAAGTCCAGGTTTGTATGAAACAGGAGATAGCAGTTCTGCTCAACAGCAAT	1277
Db	321	GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn	340
Qy	1278	AAACCTTCAGGTTCCAAACAGAGAGCTGACACTGAGTGTGAGGAGGGTCTCTGCT	1337
Db	341	LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysSerSerAla	360
Qy	1338	AAGGAGGTGAGATTGATTAATTTCCCAATTTGAGGAGGAGAGAGTGCAGTGGAAATG	1397
Db			
Db	361	LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet	380
Qy	1398	AATGATGTGAATGCCAAACGACCTGGAGTACTTCTAGCTGTAATGTAGAGTGCAGAAAT	1457
Db	381	AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn	400
Qy	1458	TCTAAGCAACATGGAGAAAGGATTTCTAAATCAGATCATTTTCATGAGAGTCCCCAAA	1517
Db	401	SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys	420
Qy	1518	GCAGAGCAAAAAAGAAACAAATGTGAATGAAACATCAAGAACAGAAAGAAAGATC	1577
Db	421	AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle	440
Qy	1578	CCTAATAACATCCACCTCACCTTCTCCAGATAAGAAATGGCTTGGAATCTCTATTGAG	1637
Db	441	ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu	460
Qy	1638	GAGATGAGAGAAATGCCAAGGTGTGGATCCGGCTGCTCCCTTGAGACCATCTGCCAAT	1697
Db	461	GluMetArgArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn	480
Qy	1698	CACAGTACTATTCGGGTAGATCTTTTGGGAATAGAGAAAGTTCCTAAACCTTTCCCA	1757
Db	481	HisThrValThrIleArgValAspLeuArgIleGlyGluValProLysProPhePro	500
Qy	1758	ACACATTTTAAAGATTTGTGGCAACAAGCATGTTAAGATGCTTGTTCAGAAACAAAC	1817
Db	501	ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn	520
Qy	1818	TTGTACCTGTGGAGATGAGATGGTGAGAGCTGCAGGACCGCGTGGGAATCTCAT	1877
Db	521	LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle	540
Qy	1878	CAGACTGCACTTCTCAACAGGCTCAGTCGGCCCCAGAACCTGAAGGATGCTATTCTGAG	1937
Db	541	GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys	560
Qy	1938	TACATGTGGCATATCTTACAAATGGACCTTACAGCTTTTACAGTTTGTATTCTTGGGATAG	1997
Db	561	TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys	580
Qy	1998	GTACTAGAAGAGCAGAGAGCTCAACACTGTATCATGCTCCATCTTCCCTGATATGGTAAA	2057
Db	581	ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys	600
Qy	2058	ATTGCACTCTGTCTGCCAATATTTGTACCCAGCAATACCACTCTCTCAAAACAGAGATG	2117
Db	601	IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet	620
Qy	2118	AATCATTTCCATCACAATGTCTACAGGAAACAGATTCGAGTTCCTTTAGCTTAATGCTTCTTC	2177
Db	621	AsnHisSerIleThrMetSerGlnGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe	640
Qy	2178	TGCAGCTTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGATTATCCAGATATT	2237
Db	641	CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle	660
Qy	2238	AACTTCAATCGGTTGTTGAAGGAGGTTTCATCAAGGAAACCCAGAGAGCTTAAACGCTC	2297
Db	661	AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu	680
Qy	2298	TTCTGCTACTTTAGAAAGTTCACAGAGAAAAAACCCACTGGGTTGGTGACATTCACAGA	2357
Db	681	PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg	700
Qy	2358	CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAAAATCTCCTGACTGCATCAT	2417
Db	701	GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis	720
Qy	2418	GTCACTTACGAGGTACCATAGAGGAAACCGCCAGGCGCATGTACAGTGTGATTTTCA	2477
Db	721	ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla	740

2478 AACGTTTCCTGAGGTGGTGAACAGTCAGAGTCTGTGCAAGAGAAATCCGCTTT 2537  
 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluGluLeuArgPhe 760  
 2538 TTAATCAACCTGAGTGTGATTTTCACGGCTCTTCACTGAGGTGCTGGATCAATGAA 2597  
 761 LeuLeuAsnProGluLeuLeuValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 2598 TGTCTTATCATCAGGTACTGACAGTACAGTGAATACACAGCTATGCCGAACATAC 2657  
 781 CysLeuIleThrGlyThrGluGlnThrSerGluThrGlyThrGlyAlaGluThrTyr 800  
 2658 CGCTGGCCCGGACCATGAAGACAGGACAAAGGACAGCTCGACAGCGCACCACT 2717  
 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgThrThr 820  
 2718 GAGATCGTCGCATCGACGCCCTCCACTTCAGACGCTACCTCGACCGATTTGTGCGCGAG 2777  
 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840  
 2778 AAGATCAGACGGGAGCTTAACAGGCTTACTGTGATTTCTTCTGCTGGAGTTTCTTCA 2837  
 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 2838 GAGACCTGTCTGACGTGCTACAGGAACTGGGGCTGTGGTCCCTTTGGGGGTGATGCT 2897  
 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
 2898 AGACTAAAGCCCTTAATACAGATCTCGCAGCTGTGCTAGTGTGACGAGACGTGTTTAT 2957  
 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 2958 TTCACCTTTGGGACCTACAGACTGATGAGACATTTACAGCATCATACATTCCTCACT 3017  
 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 3018 GAGAGAACTCAGCTGTGGAGAGTATATAAGCTGCTGCTACGATATATACATTAAGAA 3077  
 921 GluArgLysLeuThrValGlyGluValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 3078 TGCAGAACTGCTCCACCCCGGACAGACATCAAGCTTTATCCATTCATATACCATGCA 3137  
 941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 3138 GTTGAGTCTCTACACAGACACCAACCCAGCCGCGGACAAAGGACGGGGCC 3188  
 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 2  
 US-09-511-477-2  
 ; Sequence 2, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; PRIOR APPLICATION NUMBER: 2000-02-23  
 ; PRIOR FILING DATE: 09/302,812  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-511-477-2

Alignment Scores:  
 Pred. No.: 0 Length: 977  
 Score: 5184.00 Matches: 977  
 Percent Similarity: 100.00% Conservatives: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 71.58% Indels: 0  
 DB: 4 Gaps: 0

US-09-302-812-1 (1-4070) x US-09-511-477-2 (1-977)

QY 258 ATGAGTCCGGCCCGCTGTGAGCCCTGCACCAAGCGACCCGCTGGAGCGCCGCTGCA 317  
 Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
 QY 318 ACTTCTCCGGCCCGCTCGACCGCCCGAGCTTCCCGCGAGCGAGCGGCGTCTC 377  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
 QY 378 GATTCCAGGACGCTCCGGTGCAGTTCAGGTCCGCGCTCCTCTGTCAGGCTGCGCCCTG 437  
 Db 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 50  
 QY 438 GGCCGGCGGCGACAGCACCGAGGCGCGCCACCTCTCTTGTGTTTCAACACAGAGACTATA 497  
 Db 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
 QY 498 ACCAGTTGGATGGACACATAAGGAATCAAGACAGTGTGAATCAGAAAGTTTGCATAGTAAA 557  
 Db 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLeuHisSerLys 100  
 QY 558 GAAACACATACACAGAGAGATCCATGATGAGTCTGTACAAAAAGATAACCTTTAT 617  
 Db 101 GluAsnAsnAsnThrArgGluSerMetSerSerValGlnLysAspAsnPheTyr 120  
 QY 618 CAACATAACATGAAAAAATAGAAAAATGTTTCTCAGCTAGTGTGTTTGAATGACAGTT 677  
 Db 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
 QY 678 GAAAGAGTACACAGTATTGAAGCAGCATCAGCTGGGCTATGTGTAAGTGGCAGAT 737  
 Db 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
 QY 738 GAAGGGCCACACTCAGACGGCTTTGGAACTGAACTCCAGCGTAACTCTGCTATCCA 797  
 Db 161 GluGlyProHisSerSerGluArgLeuLysGluProProAlaValThrLeuValPro 180  
 QY 798 GAGCAGTTTCAGTAATGCTAATGCTCAGTCTGCCCAAGAGATGATCAGATGACACA 857  
 Db 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
 QY 858 AATAGTGGAGAGTAGAGATTAATCAGCAGTGTGTTTGCACATGTAAGCTTCGAATGCA 917  
 Db 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
 QY 918 AGCAGACCATGGAACATGAACAGAGCAGAGAGCCAGAGCCAGAGAGTGTGGCAAG 977  
 Db 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
 QY 978 GCTTGCATCTCTGCAAGACCTGTGAGGGTGTGAGGAGGAGAGACAGCTGTGTCTC 1037  
 Db 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluThrAspValValSer 260  
 QY 1038 GAGAGCCCTTGTGCGACACTCGCTCTGAGGATGTGTGCTGCTGCTGCTGCTGCTGCT 1097  
 Db 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
 QY 1098 AGATTGAATAGACAGAAAGTAGTCTAGGAAATTTCTCTCCATTTTCCAGAGAAAGTGA 1157  
 Db 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300  
 QY 1158 CTTGAGTCCACCAATGAGTGTAGATAATCCAAAAATGTTGTCAGGATTCAGAGCAGAT 1217

301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
1218 GNAGAGACAGTCCAGGTTTTCATCAACAGGAGATAGCAGTTCCTGCTCAACAGCAAAAT 1277  
321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
1278 AAACCTTCAAGGTTCCAAACAGAGAAAGCTGACCTAGTTCAGGAAAGCGTCTCTGCT 1337  
341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360  
1338 AAGGAGGTGAGATTCGATTACATTCCTCAATTTGAGGAGGAGAGAGTGCAGCTGGAATG 1397  
361 LysGlyGlyGluLeuArgLysPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380  
1398 AATGATGTGAATGCCAAACAGCTCGAAGTACTTCTAGCTGAAATGATAGAGTGCAGAAAT 1457  
381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
1458 TCTAAGCAACATGGGAGAAAGGATTCCTAAATACAGATCATTTTCATGAGTGCCTCAA 1517  
401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
1518 GCGAGGACAAAGAAAGACAACTGTAATCAAAATCAAAAGACAGAAAGAAAGATC 1577  
421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
1578 CCTAAATACATTCACCTCCTTCTCCAGATAAGAAATGGCTGGAACTCCTATTAGAG 1637  
441 ProLysThrIleProProHisLeuSerProAspLysIleThrLeuGlyThrProLysGlu 460  
1638 GAGATGAGGAGAAATCCAGGTGTGGATCCGGTCTCCTTCCCTTGAGACCATCTGCCAAT 1697  
461 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuArgProSerAlaAsn 480  
1698 CACACAGTACTATTCCGGTAGATCTTTTCGATAGGAGAGTTCCTAAACCTTTCCCA 1757  
481 HisThrValThrIleArgValAspLeuLeuArgLysGlyGluValProLysProPhePro 500  
1758 ACACATTTAAAGATTTGTGGCAACAAAGCATCTTAAGATGCTTGTTCAGAACAAAC 1817  
501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
1818 TTGTACCTCTGGAGATGAGATGTTGAGAGCTGCGAGCTGCGAGCCGCTGGAACTATT 1877  
521 LeuTrpValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
1878 CAGACTGCACTTCTCACAGCTCCTACGCGCCAGAACCTGGAAGTGTATTCTGAAG 1937  
541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
1938 TACATGTGGCATATTCTAAGAAATGGACCTTACAGCTTTGATGATTTCTGGGATAAG 1997  
561 TyrAsnValAlaLysSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
1998 GTACTAGAGAAAGCAGAGCTCAACACTGTATCAGTCCATCTTGCCTGATATGCTGAA 2057  
581 ValLeuGluGluAlaGluAlaGlnHisLeuTrpGlnSerIleLeuProAspMetValLys 600  
2058 ATTGCACCTCTGCTGCCAAATATTGTATCCAGCAATATACACCTCCTGAAACAGAGATG 2117  
601 IleAlaLeuCysLeuProAsnIleCysThrGlnProLysProLeuLeuLysGlnLysMet 620  
2118 AATCATTCATCACAATGTCCAGGAACAGATTCAGCTTTTTCAGTAAATGCTTTCTTC 2177  
621 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
2178 TGCAGCTTCCAGCAGCAATGCCAAGATGAATCAGATGATTCAGTATTCAGATATT 2237  
641 CysThrPheProArgAsnAlaLysMetLysSerGluTrpSerSerTrpProAspIle 660  
2238 AACTTCAATCGGTTGTTTGAAGAGCTTTCATCAAGAAACAGAGAGCTTAAACGCTC 2297  
661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680

2298 TTCTGCTACTTTAGAGAGTCCAGAGAAACCCACTGGGTGGTGACATTTCAAGA 2357  
681 PheCysTrpPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
2358 CAGAGCTTTGAGATTTTCCAGAGTGGGAAAGATGCTGAAAACTCCTCAGCTGATGCTAT 2417  
701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
2418 GTCACTTACGAAAGTACCATAGAGAAACCGCCAGGCGATGCTACAGTGGATTTTGA 2477  
721 ValThrTrpGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
2478 RACCGTTTCTGTTGAGGTGTAAACAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2537  
741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLysIleArgPhe 760  
2538 TTAATCAACCCCTGAGTTCATTGATTTTCAAGGCTCTTCACTGAGGTCTGATCAATGAA 2597  
761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
2598 TGTCTTATCATCAGAGTACTGACAGTACAGTGAATACACAGGCTATGCCGAAACATAC 2657  
781 CysLeuIleIleThrGlyThrGluGlnTrpSerGluTrpThrGlyThrAlaGluThrTrp 800  
2658 CGCTGGCCCGAGGCTTGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717  
801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspTrpGlnArgThrThr 820  
2718 GAGATCGTCCCATCGACGCCCTCCACTTCAGAGCTCTCTGAGCCAGTTTGTGCCGAG 2777  
821 GluIleValAlaIleAspAlaLeuHisPheArgTrpLeuAspGlnPheValProGlu 840  
2778 AGATCAGAGCGGAGCTTAAACAGGCTTACTGTGGATTTCTTCCTGCTGAGTTTCTTCA 2837  
841 LysIleArgArgGluLeuAsnLysAlaLysCysGlyPheLeuArgProGlyValSerSer 860  
2838 GAGAACTGTCTGCTGAGTGTACAGGAACTTGGGCTGTGGTCTTGGGCTGATGCT 2897  
861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880  
2898 AGACTAAAGCTTAAATACAGATCTCGGAGCTGCTGTAGCTGAGGAGGAGGAGGAGTAT 2957  
881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTrp 900  
2958 TTCACTTTGGGAGCTCAGACCTGATGAGAGACATTTACAGCATGCATACATTCCTCCT 3017  
901 PheThrPheGlyAspSerGluLeuMetArgAspIleTrpSerMetHisThrPheLeuThr 920  
3018 GAGAGGAAACTGACTGTTGGAGAAATATATAAGCTGCTGCTACGATATTAATGAAGAA 3077  
921 GluArgLysLeuThrValGlyGluValTrpLysLeuLeuArgTrpTrpAsnGluGlu 940  
3078 TGAGAAATCTGCTCCACCCCGGAGCAGACATCAAGCTTTATCCATTTCATATACATGCA 3137  
941 CysArgAsnCysSerThrProGlyProAspIleLysLeuTrpProPheIleTrpHisAla 960  
3138 GTTCAGTCTCTGTACACAGACCAACAGCCGCGGACAAAGGAGCGGGGCC 3188  
961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977

RESULT 3  
US-09-511-507-2  
; Sequence 2, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH

```

; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 2
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
US-09-511-507-2
Alignment Scores:
Pred. No.: 0 Length: 977
Score: 5184.00 Matches: 977
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 71.58% Indels: 0
DB: 4 Gaps: 0

US-09-302-812-1 (1-4070) x US-09-511-507-2 (1-977)
QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCAGCCCGCTGGAGCGCGCTGCA 317
DB 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20
QY 318 ACTTCTCCCGCGCGCTCGAGCCCGGAGCTTCCCGCGGAGGCGGAGCGCGCTCTC 377
DB 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGluArgValLeu 40
QY 378 GATTCACAGGCGCTCCGGTCAGTTCAGGTCGCGCTCTCTCTAGGTCGCGCTG 437
DB 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60
QY 438 GCGCGGGGACAGCAGCAGCGAGGAGCCACCTCTCTGTGTTTCAACAGAGACTATA 497
DB 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80
QY 498 ACCAGTTCGATGGACACTAAAGGAATCAAGACAGTTGAATCAGAAAGTTTCATAGTAAA 557
DB 81 ThrSerTrpMetAspThrLysGlyLysThrValGluSerGluSerLysHisSerLys 100
QY 558 GAAACAAACAATACAGAGAAATCCATGATGAGTCTCTGACAAAAGATAACTTTAT 617
DB 101 GluAsnAsnThrArgGluGluSerMetSerValGlnLysAspAsnPheTyr 120
QY 618 CAACATACATGGGAAAATAGAAATGTTCTCAGCTAGCTTTTGTATAGTCAACAGTT 677
DB 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140
QY 678 GAAAAGGTACACAGTATTGAAAGCAGCATCAGACTCGCGCTATGTGTAAGTGGCAGAA 737
DB 141 GluLysGlyThrGlnTyrLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160
QY 738 GAAGGCGCACACTCAGAACGGCTTTGAAAGTGAACCTCCAGCGGTAACTCTGTGTACCA 797
DB 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180
QY 798 GAGCAGTTTCAGTAAATGTCATTCAGTCAGTCGTCGCCCAAGAGGTATGATCAGAGTACACA 857
DB 181 GluGlnPheSerAsnAlaAsnValAspGlnSerProLysAspAspHisSerAspThr 200
QY 858 AATAGTAGGAGAGTAGATAAATCAGCAGTGTGTTTGCACACATGTAAAGCTTGCAGTACA 917
DB 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220
QY 918 AAGCAGATGAGATGACAGGGCAGAGAGCAGAGCCAGCCAGCAAGTGTGGCAAG 977
DB 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240
QY 978 GCTTCCCATCTGCAGAGCCCTGTGCAGGTGTCTCAGCAGGAGGACAGAGCTGTGTGCC 1037

```

---

```

DB 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260
QY 1038 GAGAGCCCTTTGTGCGGACACTGGCTCTGAGGATGTTGGTACTGGACTGAAAATCCCAAC 1097
DB 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280
QY 1098 AGATTGAATAGACAAAGTAGTCTAGGAATTTCTCTCCATTTCGAAAGAAAGTGA 1157
DB 281 ArgLeuAsnArgGlnGluSerSerLeuGlyAsnSerProProPheGluLysGluSerGlu 300
QY 1158 CCTGAGTCACCAATGGATGTAGATAATTCCAAAATAGTCTCTCAGGATTCAGAACACAGAT 1217
DB 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320
QY 1218 GAAGAGACAGTCCAGGTTTGTATGACACAGGAAGATAGCAGTTCTCTCAACAGCAAT 1277
DB 321 GluGlnThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTGACACTGAGTTGAGAGCGGCTCTCTGCT 1337
DB 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerSerAla 360
QY 1338 AAGGAGGTGAGATTCGATTACATTTCCAATTTCAAGGAGGAGAGAGTCGAGCTGGAATG 1397
DB 361 LysGlyGlyGluIleArgLeuHisPheGlnPheGluGlyGlyGluSerArgAlaGlyMet 380
QY 1398 AATGATGTGAATGCCAACACGACCTCGGAAGTACTTCTAGCCCTGAATGTAGAGTGCAGAA 1457
DB 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400
QY 1458 TCTAAGCAACATGGGAGAGAGGATTTCAAATCAAGATCAAGATCAATTTTCATGAGAGTGC 1517
DB 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420
QY 1518 GCAGAGACAAAAGAAAGAACATGTGAATGAACATCAACATCAAGAAAGAAAGAAAGATC 1577
DB 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440
QY 1578 CCTAAATACATTCACCTTCCACATTAAGAAATGGCTTGGAACTCTTATTGAG 1637
DB 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460
QY 1638 GAGATGAGGAGAAATGCCAAGGTGGATCCGGTCCGCTCCCTTGGAGACCATCTCCCAAT 1697
DB 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480
QY 1698 CACACAGTGCATTTCCGGTAGATCTTTTGGCAATAGAGAAATGCTCTTAAACCTTTCCCA 1757
DB 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500
QY 1758 ACACATTTTAAAGATTTTGGGACAAACAGCATGTTAAGATGCCTTGTTCAGACAAAAC 1817
DB 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520
QY 1818 TTGTACCTCTGGAGATGAGAAATGGTCGAGCGAGCTGCAGCAGCGCGTGGGAACTCAT 1877
DB 521 LeuTyrProValGluAspGluAsnGlyIleArgAlaAlaGlySerArgTrpGluLeuIle 540
QY 1878 CAGACTGCATCTTCAACAGGCTCACTCGGCGCCACAGAACCTGAAGGATGCTATTCTGAAG 1937
DB 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560
QY 1938 TACAATGTGGCATATTTAAGAAATGGACATTTACAGCTTTGATTTGATTTCTGGGATAAG 1997
DB 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580
QY 1998 GTACTAGAAGAGCAGAGAGCTCAACACTGTGATCAGTCCATCTTGTCCCTGATGATGTA 2057
DB 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600
QY 2058 ATTGACATCTGTCTGCCAAATATTGTACCCAGGCAATACCATCTCTCTGAAACAGAGATG 2117
DB 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620

```

```
QY 2118 AATCATTCATCACAAATGTCACAGAAAGATGCGAGTCTTTTACGTAATGCTTCTTC 2177
Db |||
QY 621 AenHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhe 640
Db |||
QY 2178 TGCAGCTTCCACGACGCAATGCCAAGATGAATCAGAGTATTCAGTATTCAGATATT 2237
Db |||
QY 641 CysThrPheProArgAsnAlaLysMetIysSerGluTyrSerSerTyrProAspile 660
Db |||
QY 2238 AACTTCATTCGGTCTTTTGAAGAGCTTCATCAAGAAACACAGAAAGCTTAAACGCTC 2297
Db |||
QY 661 AenPheAsnArgLeuPheGluGlyArgSerArgIysProGluLysLeuIysThrLeu 680
Db |||
QY 2298 TTCTGCTACTTTAGAGAGTCCACAGAGAAACCCACTGGGTGGTGACATTCACAGA 2357
Db |||
QY 681 PheCysTyrPheArgValThrGluLysIysProThrGlyLeuValThrPheThrArg 700
Db |||
QY 2358 CAGAGTCTTGAAGATTTCCAGAGTGGGAAAGATGTGAAACTCTCTGACTCGACTGCAT 2417
Db |||
QY 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720
Db |||
QY 2418 GTCACCTTACGAAGGTACCATAGAAAGAAACGCCAGGCATGCTACAGTGTGATTTGCA 2477
Db |||
QY 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740
Db |||
QY 2478 AACCGTTTCGTTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAAGAGAAATCCGCTTT 2537
Db |||
QY 741 AenArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluLeuArgPhe 760
Db |||
QY 2538 TTAATCAACCTGAGTGTGATTTTTCAGGCTCTTCACTGAGTGTCTGATCACAATGAA 2597
Db |||
QY 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780
Db |||
QY 2598 TGTCTTATCATCACAGGTACTGAGCAGTACAGTGAATACACAGCTGTGCCGAACATAC 2657
Db |||
QY 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800
Db |||
QY 2658 CGTGGGCGCGGACCATGAAGACAGGAGCGAAAGGACGACTGGCAGAGCGCGCAGCT 2717
Db |||
QY 801 ArgTrpAlaArgSerHisGluAspArgSerGluArgAspAspTrpGlnArgThrThr 820
Db |||
QY 2718 GAGATCGTCGCATCGCGCCCTCCACTTCAGAGCTACCTCGACCGAGTTGTGCCGAG 2777
Db |||
QY 821 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 840
Db |||
QY 2778 AAGATCAGACGGAGCTTAAACAGGCTTACTGTGGATTCTTCTGCTGAGATTCTTCA 2837
Db |||
QY 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860
Db |||
QY 2838 GAGAACCTGTCTGACGTGGCTACAGGAAACCTGGGCTGTGGTCTTGGGGGTGATGCT 2897
Db |||
QY 861 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 880
Db |||
QY 2898 AGACTAAAGGCTTAATACAGATCCTGSCAGCTGCTGTAGCTGAGCAGAGCTGTGTTAT 2957
Db |||
QY 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900
Db |||
QY 2958 TTCACCTTTGGGACTCAGAACTGATGAGAGACATTTTACAGCATGCATATCTCTCACT 3017
Db |||
QY 901 PheThrPheGlyAspSerGluLeuMetArgAspileTyrSerMetHisThrPheLeuThr 920
Db |||
QY 3018 GAGAGAAACTGACTGTGGAGAGATATATAGCTGTGCTGTACGATATTAATGAAGAA 3077
Db |||
QY 921 GluArgLysLeuThrValGlyGluValTyrIysLeuLeuLeuArgTyrTyrAsnGluGlu 940
Db |||
QY 3078 TGCAGAAACTGCTCCACCGCGGACACAGATCAAGCTTTATCCATTCATATACCATGCA 3137
Db |||
QY 941 CysArgAsnCysSerThrProGlyProAspileLysLeuTyrProPheIleTyrHisAla 960
Db |||
QY 3138 GTTGAGTCTGTATACAGACACCAACAGCGCGGACAAAGAGCGGGGCGC 3188
Db |||
QY 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGlyAla 977
Db |||
```

```
RESULT 4
US-09-302-812-4
; Sequence 4, Application US/09302812B
; Patent No. 6333148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1998-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-302-812-4

Alignment Scores:
Pred. No.: 0 Length: 976
Score: 4635.50 Matches: 874
Percent Similarity: 92.83% Conservative: 32
Best Local Similarity: 89.55% Mismatches: 69
Query Match: 64.01% Indels: 1
DB: 4 Gaps: 1

US-09-302-812-1 (1-4070) x US-09-302-812-4 (1-976)
QY 258 ATGAGTCCGGCCCGCTGTGAGCCCTGCACCAAGCAGCCCGCTCGGAGCGCGCTGCA 317
Db |||
QY 1 MetAsnAlaGlyProGlyCysGluProCysThrLysAlaThrArgTrpGlyAlaAlaThr 20
QY 318 ACTTCTCCGCGCGCGCTCGGAGCGCGGAGCTTCCCGGAGGAGAGCGCGCTCCTC 377
Db |||
QY 21 ThrSer---ProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeu 39
QY 378 GATTCCAAAGACCTCGGTCAGGTTCAGGTCCTCGGCTCCGCTCGTCAGGCTCGCCCTG 437
Db |||
QY 40 AspProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValPro 59
QY 438 GCGCGCGCGGACAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTATA 497
Db |||
QY 60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79
QY 498 ACCAGTTGGATGACACTAAAGGAATCAACACAGTGTGAATCAGAAAGTTTGCATAGTAAA 557
Db |||
QY 80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99
QY 558 GAAACACAAATACAGAGAGAAATCCATGATGAGTCTGTACAAAAGATAACTTTTAT 617
Db |||
QY 100 GluAsnAsnAsnThrArgIleGluSerMetSerSerValGlnLysAspAsnPheTyr 119
QY 618 CACATAACATGCAAAAATAGAAATGTTTCTCAGCTAGTGTTCATAGTCACCATGTT 677
Db |||
QY 120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139
QY 678 GAAAAAGGTACACAGTATTGAAAGCAGCATCAGCTGCGGCTATGTGTAGTGGCAGAAAT 737
Db |||
QY 140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159
QY 738 GAAGGCGCACCTCAGACAGCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db |||
QY 160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179
QY 798 GAGCAGTTCAGTAACTCTAATGTCGATCAGTCGTGCTCCCAAGAGTATCATCAGTGACACA 857
Db |||
```



180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspHisSerAspThr 199  
 858 AATAGTCAGAGAGTACAGATTAATCACAGCTTTTGGACACATGTAAAGCTTGGCAATGCA 917  
 200 AspSerGluGlnAsnArgAspAsnGlnPheLeuThrThrValIysLeuAlaAsnAla 219  
 918 AAGCAGACGATGGAAGATGAACAGCGCAGAGAACCCAGAGCCACCAAGAGTGTGGCAAG 977  
 220 LysGlnThrThrGluAspGluHisAlaArgGluAlaIysSerHisGlnLysCysSerLys 239  
 978 GCTTGCCATCTGTCAGAACCCCTGTCAGAGGTCTCAGCAGGAGGAGACAGAGCTGTGTC 1037  
 240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValPro 259  
 1038 GAGAGCCCTGTCGGACACTGCTCAGAGATGTTGGTACTGTCAGCTGAAATAATGCCAAC 1097  
 260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
 1098 AGATTGAATAGACAGAAAGTAGTCTAGAAATTCCTCCATTTGCAATTTGAAAGAAAGTCAA 1157  
 280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProPheGluLysGluSerGlu 299  
 1158 CTTGAGTCCCAATGGATGTAGATTAATCCAAATAGTCTCAGGATTCAGAGTTCAGAACAGAT 1217  
 300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
 1218 GAAGAGCAAGTCCAGGTTTTCATCAACAGAGAGATAGCAGTCTCTGCTCAAAACAGCAAT 1277  
 320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsn 339  
 1278 AAACCTTCAAGGTTCCAAACAGAGAGTACACTGATGTAGAGAGCGGTCCTCTGCT 1337  
 340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
 1338 AAGGAGGTGAGATTCGATTAATTCATTTCCATTTGAGGAGGAGAGTCCGACTGCAATG 1397  
 360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
 1398 AATGATGTGAATGCCAAACGACTCGAAGTACTTCTAGCTCGAATGTAGAGTGCAGAAAT 1457  
 380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
 1458 TCTAAGCAACATGGGAGAAAGATTTCAAAATCACAGATCAATTCATGAGTGGCCCAA 1517  
 400 SerLysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
 1518 GCAGAGGCAAAAGAAAGAAACAATGTGAATGAAATCAAAAGACAGAAAGAGATC 1577  
 420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439  
 1578 CCTAAATACATTCACCTTCCTCCAGATPAGAAATGGCTTGGAACTCTTATTGAG 1637  
 440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProLysGlu 459  
 1638 GAGATGAGGAGATGCCAGGTGTGGATCCGCTGCTCCCTTGAGCATCTGCCAAT 1697  
 460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
 1698 CACACAGTGAATTCGGGTAGATCTTTTCGAATAGGAGAGTCTCTCAACCTTTCCCA 1757  
 480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
 1758 ACATTTTAAAGATTTGGGCAACAACAGCATGTTAAGATGCTTGTTCAGAACAAAC 1817  
 500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
 1818 TTGTACCTCTGGAGATGAGATGTGTGAGGAGCTCAGCAGCCCGGTGGGAACTCAT 1877  
 520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
 1878 CAGATGCACTTCTCAACAGCTCCTCGGCCAGAACCTTGAAGGATGCTATTCTCAAG 1937  
 540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 559

1938 TACAATGTGGCATATTCTAAGAAATGGACTTTTACAGCTTTGATTGATTTCTGGATAAG 1997  
 560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
 1998 GTACTAGAAAGCAGAGCTCAACACTTGTATCAGTCCATCTTGTCTGTATATGTGTA 2057  
 580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
 2058 ATTGACATCTGTCTGCCAAATATTGTACCAGCAATACACTCTCTCCCTGAAACAGAAATG 2117  
 600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
 2118 AATCATTTCCATCACAATGTACAGAAACAGATTCCCACTCTTTAGCTTAATCTCTTC 2177  
 620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
 2178 TGCACTGTTTCCACGACCAATTCGCAAGATGAAATCAGAGTATTCAGTATTCAGATATT 2237  
 640 CysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
 2238 AACTTCATCTGTTGTTGAAGAGCTTCATCAAGGAAACCCAGAGAGCTTAAACAGCTC 2297  
 660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
 2298 TTCTGCTACTTTAGAGAGTCAACAGAGAAAAAACCCACTGGGTGGTGACATTTCAAGA 2357  
 680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
 2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAAGATGTGAAAACTCTCAGTCAGCTGCAT 2417  
 700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
 2418 GTCACTTACGAGGTACCATAGAGGAAACCGCCAGGCGCATCTACAGTGGATTTTGA 2477  
 720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
 2478 AACCTGTTTCTGTTGAGGTGGTGAACCACTGAGGACTTGTGCAAGAAAGAAATCCGCTT 2537  
 740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 759  
 2538 TTAATCAACCTCTGATGATTTTTCAGGCTCTTCTCAGTGGTCTGATCACATGAA 2597  
 760 LeuIleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
 2598 TGTCTTATCATCACAGGTACTCAGCAGTACAGTGAATACACAGGCTATCCGCAAAACATAC 2657  
 780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799  
 2658 CGCTGGCCCGCAGCCATGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717  
 800 ArgTrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgArgCysThr 819  
 2718 GAGATGTCGCCATGACGCCCTCCACTTCAGACGCTACTCCAGCCAGTTGTGGCCCGAG 2777  
 820 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 839  
 2778 AAGATCAGAGCGGAGCTTAAAGGCTTACTGTGGATTTCTCGTCTCGAGTTTCTTCA 2837  
 840 LysMetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 859  
 2838 GAGAACCTGTCGTCAGTGGCTTACAGGAAACTCGGGCTGTGGTCCCTTGGGGGTGATGCT 2897  
 860 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 879  
 2898 AGACTAAAGCCCTTAATACAGATCTCGGAGCTGTGTAGCTGAGCAGGAGACGTGGTTAT 2957  
 880 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyr 899  
 2958 TTCACCTTTGGGAGCTCAGAACCTGATGAGACATTTTACAGCATGATCATTTCTCTCACT 3017  
 900 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThr 919







60 GlyGlnAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 79  
498 ACCAGTTGGATGACACATAAAGGAATCAAGACAGTGAATCAGAAAGTTGGCATAGTAAA 557  
80 ThrSerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLys 99  
558 GAAACACACATACACAGAGAGATCCATGATGATGTTCTGACAAAAGATATACTTTTAT 617  
100 GluAsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAspPheTyr 119  
618 CAACATAACATGGAATAATTAGAAAATTTCTCAGCTAGCTTTTGATAAGTCACCAAGTT 677  
120 GlnHisAsnValGluLysLeuValAsnValSerGlnLeuSerLeuAspLysSerLeuThr 139  
678 GAAAGAGGTACACAGTATTGAACAGCAGATCAGACTCGGGTATGTGTAGTGCAGCAAT 737  
140 GluLysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 159  
738 GAAGGCCACACTCAGAACGGCTTTTGGAAAGTCAACCTCCAGCGGTAACTCTGGTACCA 797  
160 GluGlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValPro 179  
798 GAGCAGTTTCAGTAAATGTCGATCGATCGTCCCAAGGATGATCAGATGACACACA 857  
180 GluGlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThr 199  
858 AATAGTCAGGAGATGAGATAATCAGCAGTTTTTGACATGATAAGCTTGCAGTGCACA 917  
200 AspSerGluGluAsnArgAspAsnGlnPheLeuThrValLysLeuAlaAsnAla 219  
918 AAGCAGACGATGGAGATGAACAGCGGACAGAGACCCAGAGAGTGTGGCAAG 977  
220 LysGlnThrThrGluAspGluHisAlaArgGluAlaLysSerHisGlnLysCysSerLys 239  
978 GCTTGCCTCCTCGCAGAGCCTGTGCGAGGTGTGACAGGAGAGACAGACGCTGGTPTCC 1037  
240 SerCysHisProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValPro 259  
1038 GAGAGCCCTGTGCGACACTGGCTCTGAGATGTGTGATGCTGACGATCAAAAAGTCCAC 1097  
260 LysSerProLeuSerAspValGlySerGluAspValGlyThrGlySerLysAsnAspAsn 279  
1098 AGATTGAATAGACAAAGAAAGTAGTCTAGAAATTCCTCCATTTGAGAAAGAAAGTCAA 1157  
280 LysLeuIleArgGlnGluSerCysLeuGlyAsnSerProProPheGluLysGluSerGlu 299  
1158 CCTCAGTCCCAATCGATGTAGATAATTCCAAAATAGTTCTCAGGATTCAGAGCAGAT 1217  
300 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 319  
1218 GAAGAGCAGACTCCAGGTTTGTGATGACAGGAGATAGCAGTTCTGCTCAACACAGCAAT 1277  
320 GluGluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsn 339  
1278 AAACCTTCAAGGTTCCAAACAGAGAGCTGACACTGAGTTGAGGAACGGTCTCTGCT 1337  
340 LysProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThr 359  
1338 AAGGAGGTGAGATTGCAATTACATTTCCAAATTTGAAGGAGAGAGATGCTGAGTGGAAATG 1397  
360 LysGlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMet 379  
1398 AATGATGTGAATGCCAAGACGCTGGAAGTCTTCTAGCTGAATGATGATGACAGAAAT 1457  
380 AsnAspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsn 399  
1458 TCTAAGCAACATGGAGAAAGGATTCTAAATATCAGATCATTTTTCATGAGATGCCCAAA 1517  
400 SerLysGlnHisGlyLysAspSerLysIleThrAspHisLeuMetArgLeuProLys 419  
1518 GCAGAGGACAAAGAAAAGAACATGTGAAATGAAACATCAAGACAGAAAGGAGATC 1577  
420 AlaGluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIle 439

1578 CCTAAATACATTCACCTCCACTCCCTTTCTCCAGATAAGAAATGGCTTGGAATCCTATTGAG 1637  
440 ProLysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 459  
1638 GAGATCAGGAGATGCAAGAGTGTGGATCCGCTGCTCCCTCCCTGAGACCATCTGCCAAT 1697  
460 GluMetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsn 479  
1698 CACACAGTACTATTCCGGTAGATCTTTTCCGATAGTAGAGAGTCTCTTAACCTTCCCA 1757  
480 HisThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 499  
1758 ACACATTTTAAAGATTGTGGGACAAACAGCATGTTAAAGATGCTGTTTTCAGAACAAAC 1817  
500 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 519  
1818 TTCTACCTCTGGAAGATGAGATGTGAGCGAGCTGACGAGCGCGGTGGGAATCTCAT 1877  
520 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 539  
1878 CAGACTGCCTCTCAACAGGCTCACTCGGCCCCAGAACCTGAGAGGATGCTATTCTCAAG 1937  
540 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLys 559  
1938 TACAATGTGCATATTTCTAAGAAATGGGACTTTTACAGCTTTTGATTGATTCTCGGATAAG 1997  
560 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 579  
1998 GTACTAGAGAAAGCAGAGCTCAACACTTGTATCAGTCCATCTGCTGATATGGTAAA 2057  
580 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 599  
2058 ATTGCATCTGTCTGCCAATATTTGACCAGCAATACACTCTCTGAAACAGAAAGATG 2117  
600 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 619  
2118 AATCATTTCCATCAATGTACAGAGAACAGATTGCCAGTCTTTTGTAGTAAGTCTTCTTC 2177  
620 AsnHisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 639  
2178 TGACGTTTCCAGCGCAATGCCAGATCAATCAAGATTCAGATTCAGATTCAGATATT 2237  
640 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 659  
2238 AACTTCATCGTGTGTGTGAAGCAGCTTCATCAAGGAAACACAGAGAGCTTAAACGCTC 2297  
660 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 679  
2298 TTCTGCTACTTTAAGAGAGTCACAGAGAAAACCCACTGGTGTGGTGCATTCACAGAA 2357  
680 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 699  
2358 CAGAGTCTTGAAGATTTTCCAGAGTGGGAAGATGTCAAAAACCTCCGACTCCGACTGCAT 2417  
700 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 719  
2418 GTCACTTTACGAAGGTACCATAGAGGAAACGGCCAGGCGCATGCTACAGGTGGATTTTGA 2477  
720 ValThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAla 739  
2478 AACCGTTTCCGTGGAGGTGGTGAACCAAGTCCAGGACTTGTGCAAGAGAAATCCGCTTT 2537  
740 AsnArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluIleArgPhe 759  
2538 TTAATCAACCTTGATTTGATTTTCCAGCTCTTCACTGAGGTGCTGGATCACAATGAA 2597  
760 LeuIleAsnProGluLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGlu 779  
2598 TGTCTTATCATCAGAGTACTGAGCAGTACAGTGAATACACAGGCTATGCCAAACATAC 2657  
780 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 799



2478	QY	AA	CCGTTCTCGTTGGAGGTGGTGTAAACAGATGCAGAGCACTTGTGCAGAGAAATCCCGTTT	2539
732	Db	Asn	ArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe	751
2538	QY	TTA	ATCAACCCCTCAGTTGATGTTTTCACGCCTCTTCACTCAGGTGCTGGATCACAATGAA	2599
752	Db	Leu	IleAsnProGlnLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu	771
2598	QY	TGCT	TATCATCACAGTACTGAGCAGTAGTACAGTGAATATACACAGGCTATGCCGAAACATAC	2659
772	Db	Cys	LeuIleIleThrGlyThrGluGlnTyrSerGlnTyrThrGlyTyrAlaGluThrTyr	791
2658	QY	CGCT	GGCCCGGAGCCATGNAGACAGAGCGGAAGGACGACACTGGCAGAGCGCAGCACT	2717
792	Db	Arg	TrpAlaArgSerHisGluAspGlySerGlnIysAspAspTrpGlnArgArgCysThr	811
2718	QY	GAGAT	CGTCCCATCGACGCGCTCCACTTCAGACGCTACCTCGACCACTTTGTGCCCGAG	2777
812	Db	Glu	IleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu	831
2778	QY	AGA	TACAGCGGAGCTTAACAAGCTTACTGTGGATTCTTCGTCCTGAGATTCTTCA	2837
832	Db	Lys	ValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer	851
2838	QY	GAGA	ACTGCTGCTCAGTGGCTACAGGAAACTGGGGCTGTGGTGCTTTGGGGGTGATGCT	2897
852	Db	Glu	AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla	871
2898	QY	AGACT	AAAGCCTTAATACAGATCTCTGGCAGCTGCTAGCTGACGAGACGCTGTTAT	2957
872	Db	Arg	LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr	891
2958	QY	TTCA	CCCTTTGGGACTCAGACTCATCAGAGACATTTACAGCATGTCATACATTCCTCACT	3017
892	Db	Phe	ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr	911
3018	QY	GAGA	GAGAACTGACTGTTTGGAGAGTATATAAGCTGCTGCTTACGATATTACAATGAAGA	3077
912	Db	Glu	ArgLysLeuAspValGlyLysValTyrLysLeuLeuAspTyrTyrAsnGluGlu	931
3078	QY	TGC	AGAACTGCTCCACCCCGGACCAACATCAAGCTTTTATTCATTATCATATACCATGCA	3137
932	Db	Cys	ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla	951
3138	QY	GTT	CAGTCTCTGTACACAGACCAACCAACGCGCGGACAAAGGACGGGG	3185
952	Db	Val	GluSerSerAlaGluThrThrAsnMetProGlyGlnLysAlaGly	967

## RESULT 8

```

US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING
; TITLE OF INVENTION: THE PROTEINS
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAID 201
; CURRENT APPLICATION NUMBER: US/09/09-511-477-6
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

```

## Alignment Scores:

```

Pred. No.:      0      Length:      968
Score:          4332.50  Matches:      820
Percent Similarity: 89.86%  Conservative: 57
Best Local Similarity: 84.02%  Mismatches: 90
Query Match:      59.82%  Indels: 9
DB:              4      Gaps: 5

US-09-302-812-1 (1-4070) x US-09-511-477-6 (1-968)

QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCAGCCCGCTGGAGCGCCCTGCA 317
Db 1 MetSerAlaGlyProGlyTriPgluProCysThrLys---AlaArgTriPgluAlaGly 19
QY 318 ACTTCTCCGCGCGCTCCGAGCCCGCGAGCTCCCGCGCAGCAGCGCGCTCCTC 377
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39
QY 378 GATTCACAGGAGCTCCCGTGCAGTTCAGGGTCCCGCGCTCCTCCTCAGGCTGCCCGCTG 437
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
QY 438 GCGCGGCGGACAGACCGAGCGAGCGCCACTCTCTGTGTTTCAACAGAGACTATA 497
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79
QY 498 ACCAGTTCGATGGACACTAAAGGATCAAGCAGTTGAATCAGAAAGTTGTCATAGTAA 557
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96
QY 558 GAAACAAACATACAGAGAAATCCATGATGAGTCTCTGTACAAAAAGATAACTTTAT 617
Db 97 GluAsnAsnAsnThrArgIleAspSerMetMetSerSerValGlnLysAspAsnPheTyr 116
QY 618 CAACATAACATGGAAATATAGAAATGTTCTCAGTAGTGGTTTGTAAAGTCACAGTT 677
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136
QY 678 GAAAGAGGTACACAGTATTGAGCAGCATCAGCTCGCGGTATGTGTAAAGTGGCAGAT 737
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156
QY 738 GAAGGGCCACACTCAGACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176
QY 798 GAGCAGTTCAGTAATGCTGATCAGTCTGCTGCCCAAGGATGATCAGAGTCACACA 857
Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196
QY 858 AATAGTGGAGGAGTAGAGATATACAGCATGTTTGGACATGTAAAGCTTCGGAATGCA 917
Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216
QY 918 AAGCAGACCATGCAAGATGAAACAGGCGCAGAGAGCCAGCAGCAGAGTGTGGCAAG 977
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233
QY 978 GCTTGCATCTCGACAGCCCTGTGAGGCTGTGAGGAGTGTGGTACTGGAATAATGCCAAC 1037
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 253
QY 1038 GAGAGCCCTTGTGCGACACTGCTCTGAGGAGTGTGGTACTGGAATAATGCCAAC 1097
Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273
QY 1098 AGATTGAATAGACAGAAAGTAGTCTAGGAAATCTCTCCATTGTGAGAAAGAAAGTGA 1157
Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 293
QY 1158 CTTGAGTCAACATGATGATAGATAATTCAAAATAGTTGTCAGGATTCAGAGCAGAT 1217
Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313

```

```

QY 1218 GAAGAGCAAGTCCAGGTTTTGTAGTAAACAGGAAGATAGTCTGTCTCAACACAGCAAT 1277
Db 314 GluGluThrSerProValPheAspGluGlnAspArgSerSer---GlnThrAlaAsn 332
QY 1278 AAACCTTCAAGGTTCCAAACCAAGAGAGCTCAGTCTGAGTTGAGGAAACGGTCTCTGCT 1337
Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgGlyLeuThr 352
QY 1338 AAGGAGGTGAGATTCGNTTACATTTCAATTTGAAGAGGAGGAGAGTCCGAGTCCGAATG 1397
Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAlaAlaGlyThr 371
QY 1398 AATGATGTAATGCCAAACGACCTCGGAAGTACTTCTAGCTGAATGTAGAGTGCAGAAAT 1457
Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391
QY 1458 TCTAAGCAACATGGGAGAAAGGATTTCTAAATCAAGATCATTTTCATGAGAGTGCCTAAA 1517
Db 392 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysSerLys 411
QY 1518 GCAGAGGCAAAAGAAAGAACATGTCAAATGAAACATCAAGAACAAAGAGGAGATC 1577
Db 412 SerGluAspArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431
QY 1578 CTTAAATACATTCACCTCACCTTCTCCAGATAGAAATGCTTGGAACTCTTATTGAG 1637
Db 432 ProLysTyrIleProProAsnLeuProGluLysLysTrpLeuGlyThrProIleGlu 451
QY 1638 GAGATGAGGAAATGCCAAGGTGGGATCCGGTCCCTCCCTTGGAGACCATCTCCCAAT 1697
Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471
QY 1698 CACAGTGCATATTCGGGTAGATCTTTTGGCAATAGCAGAGATTCCTTAAACTTTTCCA 1757
Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491
QY 1758 ACACATTTTAAAGATTTTGTGGCAACAAAGCATGTTAAGATGCCCTTGTTCAGAACAAAC 1817
Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511
QY 1818 TTGTACCTCTGGAGAGATGAGAAATGGTGGAGCGCTGAGGAGCGCGGTGGGAATCATTT 1877
Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531
QY 1878 CAGACTGCATTTCAACAGGCTCACTCGGCCCAAGAACCTGAAGATGCTATTCTGAAG 1937
Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaLeuLys 551
QY 1938 TACATGTGGCATATTTCAAGAAATGGGACTTTTACAGCTTTGATTTCTGGGATAG 1997
Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571
QY 1998 GTACTAGAGAGCAGAGCTCAACACTTGTATCAGTCCATCTTCCTGATATGCTGAAA 2057
Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591
QY 2058 ATTGCACCTCTGTGCGCAATATTTGTACCAGCCCAATACCACTCCTGAAACAGAGATG 2117
Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 611
QY 2118 AATCATTCATCACAATGTCACAGAAACAGATGCGCATCTTTAGCTTAATGCTTTCTTC 2177
Db 612 AsnHisSerValThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631
QY 2178 TGCAGTTTCCACGACGCAATGCCAAGATGAAATCAGAGTATTCAGTATTCACATATT 2237
Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerLysTrpProAspLys 651
QY 2238 AACTTCAATCGGTTGTTTGAAGGAGCTTCATCAAGGAAACCAAGAGAGCTTAAACAGCTC 2297
Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671

```

```

2298 TTCTGCTACTTTAGAGAGTACAGAGAAAAACCCACTGGGTTGGTGCATTCACAGA 2357
672 PheCysTy-PheArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 691
2358 CAGAGCTTGAAGATTTCCAGAGTGGGAAGATGTGAAATCTCTGACTCGACTGCAT 2417
692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711
2418 GTCACTTACGAAGTACCATAGAGAAACCGCCAGCGCATGTCTACAGTGGATTTGCA 2477
712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731
2478 AACCGTTTCCTGGAGTGTGTAAACAGTGCAGGACTGTGTGCAAGAGAAATCCGCTTT 2537
732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751
2538 TTAATCAACCTCAGTTGATTGTTTACCGCTCTTCACTCAGGTGTGTGATCAATGAA 2597
752 LeuLeuAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771
2598 TGTCTTATCATCAGGTACTGACGAGTACAGTGAATACACAGCTATGCCGAAACATAC 2657
772 CysLeuIleIleThrGlyThrGluGlnTyr-SerGluTyrThrGlyTyrAlaGluThrTyr 791
2658 CGCTGGCCCGGAGCCATGAAGCAGAGCGAAGGACGACCTGGCAGAGCGCCACGACT 2717
792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArg-CysThr 811
2718 GAGATCGTCCCATCGACGCCCTCCACTTCAGACGCTTACCTCGACAGTGTGTGCCGAG 2777
812 GluIleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGlu 831
2778 AAGATCAGCGGAGCTTAAACAAGCTTACTGTGATTTCTGCTCGTGGATTTCTTCA 2837
832 LysValArgArgGluLeuAsnLysAlaTyr-CysGlyPheLeuArgProGlyValProSer 851
2838 GAGAACCTGTCTCAGTGGCTACAGGAACTGGGCTGTGGTGTGGTGGGTTGATGCT 2897
852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871
2898 AGACTAAAGCTTAAATACAGATCCTCGCAGCTCTGTAGCTGACGAGACGTGGTTTAT 2957
872 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaAlaGluArgAspValValTyr 891
2958 TTCACTTTGGGACTCAGAACTGATGAGACATTTACAGCATGATCATCTTCTCACT 3017
892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911
3018 GAGAGGAACCTGACTGTGGAGAGTATATAGCTGTCTGCTTACCATATTACAAATGAAGAA 3077
912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931
3078 TGCAGAAATGCTCTCCACCCCGGACACAGATCAAGCTTTATCATTCATATACCATGCA 3137
932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951
3138 GTTGAGTCTCTGATCAGACACCAACCGCCGGGACAAAGGACGGGG 3185
952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGly 967

```

RESULT 9

```

US-09-511-507-6
; Sequence 6, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201

```

```

; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
US-09-511-507-6

Alignment Scores:
Pred. No.: 0 Length: 968
Score: 4332.50 Matches: 820
Percent Similarity: 89.86% Conservative: 57
Best Local Similarity: 84.02% Mismatches: 90
Query Match: 59.82% Indels: 9
DB: Gaps: 5

US-09-302-812-1 (1-4070) x US-09-511-507-6 (1-968)
QY 258 ATGAGTGGGGCCCGCTGTGAGCCCTGCACCAAGCGACCCCGCTGGAGCGCGCTGCA 317
Db 1 MetSerAlaGlyProGlyTrpGluProCysThrLys---AlaArgTrpGlyAlaAlaGly 19
QY 318 ACTTCTCCCGCCCGCTCGGACGCCCGGAGCTTCCCGGAGGAGGAGCGCGCTCCTC 377
Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgArgValLeu 39
QY 378 GATTTCACAGACGCTCCGCTGAGGTTCCAGGCTCCCGCTCTCTCGTCAGGCTCGCCCTG 437
Db 40 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 59
QY 438 GCGCGCGCGGACAGACACCGAGGACGCCACCTCTCTGTTTTCACAGAGACTATA 497
Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79
QY 498 ACCAGTTGGATGACACTAAAGGAATCAAGACAGTGTGAATCAGAAAGTTTGCATAGTAA 557
Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96
QY 558 GAAACCAACAATAACAAGAGAAATCCATGATCAGTGTGTACAAAAGATATACTTTAT 617
Db 97 GluAsnAsnAsnThrArgIleAspSerMetSerSerValGlnLysAspAsnPheTyr 116
QY 618 CAACATACATGAAATAATAGAAATAGTTCTCAGCTAGTGTTCATAGTCCACAGTT 677
Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136
QY 678 GAAAGAGGTACACAGATATTGAAAGCAGCATCAGACTGCGGCTATGTAAAGTGCAGAA 737
Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTrpGlnAsn 156
QY 738 GAAGGGCCACACTCAGAACCGCTTTTGGAAAGTGAACCTCCAGCGGTAACTCTGGTACCA 797
Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176
QY 798 GAGCAGTTCTAGTAAATGCTAATGTCGATCGTCTGCCCAAGGATGATCAGTGCACACA 857
Db 177 LysGlnLeuSerSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196
QY 858 AATAGTCAGGAGAGTACAGATAATCAGCAGTGTTCACATGTGTAAAGCTTGGCAATGCA 917
Db 197 AspHisGluGluAspArgAspAsnGlnPheLeuThrProIleLysLeuAlaAsnThr 216
QY 918 AAGCAGACGATGGAAGATGAACAGGGCGAGAGAGCCAGAACCCACAGAGTGTGGCAAG 977
Db 217 LysProThrValGlyAspGlyGln-----AlaArgSerAsnCysLysCysSerGly 233
QY 978 GCTTGCATCTCTGCAGAGCTGTGAGGTTGTGCAGGAGGAGGAGACAGACGCTGGTGTCC 1037
Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253

```







```

US-09-302-812-8
; Sequence 8, Application US/09302812B
; Patent No. 633148
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/302,812B
; CURRENT FILING DATE: 1999-04-30
; EARLIER APPLICATION NUMBER: 60/083,768
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
US-09-302-812-8

Alignment Scores:
Pred. No.: 2,46e-93 Length: 768
Score: 1063.50 Matches: 244
Percent Similarity: 57.31% Conservative: 93
Best Local Similarity: 41.50% Mismatches: 198
Query Match: 14.69% Indels: 53
DB: 4 Gaps: 15

US-09-302-812-1 (1-4070) x US-09-302-812-8 (1-768)
QY 1506 AGAGTCCCAAGACAGAGGCAAGAAAGCAATGTAATGAACATCAAGAACCA 1565
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 45 ArgMetSerLysSerProAspGlyGlyLeuSerGluLeuGluGluGluGluGlu 62
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1566 GAAAGGAAGATCCCTAAATACATTCACCTCACCTT-----TCTCAGATAAGAAATGG 1619
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 63 -----ProGluAsnLeuAlaAsnSerLeuAspSerTrp 74
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1620 CTGTGGAATCTCTATTGAGAGATG-----AGGAGATGCCAGGTGTGGATCCGGCTG 1673
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 75 ArgGlyValSerMetGluAlaAlaHisArgAsnArgGlnProPheGluLeuGluAsnLeu 94
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1674 CCTCCCTTGAGACCATCTGCCAATCACACAGTACTTTCGGGTAGATCTTTTCGGAATA 1733
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 95 ProProValThrAlaGlyAsnLeuHisArgValMetTyrGln-----LeuProIle 111
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1734 GGAGAAGTTCCTAAACCTTTCCCAACACATTTTAAAGATTGTGGGACACAAAGCATGTT 1793
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 112 ArgGluThrPro-----ProArgProTyrLysSerProGlyLysTrpAspSerGluHisVal 130
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1794 AAGATGCTTCTCAGACACAAACTTGTACCTGTGGAGATGAGATGTTGGAGCGAGCT 1853
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 131 ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr 150
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1854 GCAGGCGCGGTGGAACTCATTCAGACTGCTCTCAACAGGCTCACTCGGCCCCAG 1913
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 151 IleAspPheArgTrpGluMetIleGluArgAlaLeuLeuGlnProIleThrCysGlu 170
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1914 AACCTGAAGATGCTATTCTGAAGTACAAATGTGGCATATTTAGAAATGGACATTTACA 1973
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 171 GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTrpHisPheArg 190
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1974 GCTTTGATGATTTCTGGATAGGTACTAGNAGNAGCAGAGCTCAACACTTGTATCAG 2033
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 AlaLeuHisGlnLeuLeuAspGluLeuAspGluSerGluThrArgValIlePheGlu 210
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2034 TCCATCTTGCTGATATGTGTAATATGCTCTCTGCGCAATATTTGTACCCAGCCA 2093
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 211 AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro 230
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 2094 ATACCACTCTCTGAAACAGAGATGAATCATTCATCCATCACAATGTCACAGAAACAGATGCC 2153
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 231 ValProLeuLeuLysHisIleLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer 250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2154 AGTCTTTAGCTAAATGCTTTCTTCGACGTTTCCACGACGCAATGCC---AAGATGAAA 2210
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 251 CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys 270
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2211 TCAGAGTATTCAGTATTCAGATATTAATCTCAATCGGTTCTTTGAAGAGCGTTCATCA 2270
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 271 SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro 290
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2271 AGGAAACCAAGAGAGCTTAAACGGCTCTCTCTCTACTTACTTAGAGAGTC-----ACAGAG 2324
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 291 AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu 310
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2325 AAAAAA-----CCCACTGGTTGGTGCATTCACAAAGACAGAGT-----CTT 2366
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 311 ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro 330
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2367 GAAGATTTTCCAGAGTGGGAAAGATGTGAAAACCTCTCTG-----ACTCGACTGCATGTC 2420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 331 GluHisLeuIleAspTrpSerGlnSerAlaAlaProLeuGlyAspValProLeuHisVal 350
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2421 ACTTACGAAGTACCATAGAGAAAGCCGCGCATCTACAGCTGGATTTTGCAAAAC 2480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 351 AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn 370
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2481 CGTTTCGTTGGAGCTGTGTAAACAGTCAGGAGTGTGTGCAAGAAAGAAATCCGCTTTTA 2540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluIleArgPheVal 390
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2541 ATCAACCTCGAGTGTGTTTTCACGCTCTTCACTGAGTGTCTGATCACAATGATGT 2600
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 391 IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuArgPropheGluAla 410
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2601 CTATCATCACAGTACTCAGCAGTACAGTGAATACACAGGCTATCGCGAAACATACCCG 2660
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 411 LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu 430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2661 TGGCGCCGAGCCATGAAGACAGAGCAGGAAAGGACGACTGGCAGAGGCGCAGACTGAG 2720
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 TrpSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla 450
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2721 ATCGTGCATCATCAGCGCTTCCATCTCAGACGCTACTCGACCGATTGTGCCGAGAAG 2780
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 451 IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisGlnTyrArgGluAspLeu 470
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2781 ATCAGACGGGAGCTTAACAGGCTTACTGTGGATTCTTCTCGT----- 2822
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 471 MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTrpMetValThrProPro 490
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2823 CCTGGAGTCTTTCCAGAGAACCTGTCTGCGTGTGCTGCTACAGAACTGGGCTGTGGTGC 2882
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2883 TTTGGGGGTGATGCTAGACTAAAGCTTTAATACAGATCTCTGGCAGCTGTGTAGCTGAG 2942
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2943 CGAGACTGGTGTATTTCCTCTTGGGACTCTCAGAACTGATGAGAGACTTTACAGCATG 3002
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 523 ArgProLeuAlaTyrTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3003 CATACATTCCTCTCAGAGAGAACTGACTGTGTGGAGAACTATATAAGCTGTGCTACGA 3062
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 543 TrpLeuLeuPheArgAsnAspGlyThrThrValGlnGlnLeuTrpSer---IleLeuArg 561
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3063 TATTACAAATGAGATCAGAAACTGCTCCACCCCGACCA----- 3104
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 SerTyrSerArgLeuIleLysGluLysSerSerLysGluProArgGluAsnLysAlaSer 581
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



131	ArgLeuProCysAlaProGluSerLysTyrProArgGluAsnProAspGlySerThrThr	150
1854	GCAGGCAGCGGTGGGAACCTCATTACAGCTGCACCTTCTCAACAGGCTCACTCGGCCCGC	1913
151	IleAspPheArgTyrGlnMetIleGluArgAlaLeuLeuGlnProIleLysThrCysGlu	170
1914	AACCTGAAGAGTCTATTCTGAAGTACAATGTGGCATATTCTAAGAAATGGGACTTTACA	1973
171	GluLeuGlnAlaAlaIleIleSerTyrAsnThrThrTyrArgAspGlnTyrHisPheArg	190
1974	GCTTTGATTGATTTCTGGGATAAGGTACTAGAGAAGCAGCAAGCTCAACACTTGATCAG	2033
191	AlaLeuHisGlnLeuLeuAspGluGluLeuAspGluSerGluThrArgValPhePheGlu	210
2034	TCCATCTTCCTGATATGGTGAATAATGCACCTCTGCTGCAATATTTGTACCCAGCCCA	2093
211	AspLeuLeuProArgIleIleArgLeuAlaLeuArgLeuProAspLeuIleGlnSerPro	230
2094	ATACCACTCCTGAAACAGAGATGAATCATTCATCACAAATGTCACAGAACAGATTGCC	2153
231	ValProLeuLeuLysHisHisLysAsnAlaSerLeuSerLeuSerGlnGlnIleSer	250
2154	AGTCTTTTACCTAATGCTTTCTTCTCGACGTTTCCAGGCGCAATGCC---AAGATGAAA	2210
251	CysLeuLeuAlaAsnAlaPheLeuCysThrPheProArgArgAsnThrLeuLysArgLys	270
2211	TCAGAGTATTCAGGTATCCAGATATTAATCTCAATCGGTGTGTGAAGCAGCTTCATCA	2270
271	SerGluTyrSerThrPheProAspIleAsnPheAsnArgLeuTyrGlnSerThrGlyPro	290
2271	AGGAACCCAGAGAGCTTAAACCGCTTCTTGCTACTTTAGAAGAGTCT-----ACGAG	2324
291	AlaValLeuGluLysLeuLysCysIleMetHisTyrPheArgArgValCysProThrGlu	310
2325	AAAAA-----CCCATCTGGTGGTGCATTCACATTCACAGACAGAGT-----CTT	2366
311	ArgAspAlaSerAsnValProThrGlyValValThrPheValArgArgSerGlyLeuPro	330
2367	GAAGATTTCCAGAGTGGGAAGATCTGAAAACTCCTG-----ACTCCAGCTGCATGTC	2420
331	GluHisLeuIleAspTyrPserGlnSerAlaAlaProLeuGlyAspValProLeuHisVal	350
2421	ACTTACAGAGGTACCATAGAGAAACGGCCAGGCATGTCTACAGTGGATTTTGCAAC	2480
351	AspAlaGluGlyThrIleGluAspGluGlyIleGlyLeuLeuGlnValAspPheAlaAsn	370
2481	CGTTTCGTGGAGTGGTGAACCAAGTCGAGGACTTGTGCAAGAAGAAATCCGCTTTTAA	2540
371	LysTyrLeuGlyGlyGlyValLeuGlyHisGlyCysValGlnGluGluIleArgPheVal	390
2541	ATCAACCCCTCAGTTGATTTTCCAGCTCTTCACTGAGTGTCTGGATCACAATGAATGT	2600
391	IleCysProGluLeuLeuValGlyLysLeuPheThrGluCysLeuA-gProPheGluAla	410
2601	CTTATCATCACAGTACTGACGAGTACAGTGAATATACACAGGCTATGCGGAACATACCGC	2660
411	LeuValMetLeuGlyAlaGluArgTyrSerAsnTyrThrGlyTyrAlaGlySerPheGlu	430
2661	TGGCCCGGAGCCATGAAGACAGAGCGAAGGAGCAGCTGGCAGAGCGGCACCACTGAG	2720
431	TyrSerGlyAsnPheGluAspSerThrProArgAspSerSerGlyArgArgGlnThrAla	450
2721	ATCGTCGCCATCAGCGCCTCCACTTCCAGCGTACTCTGACCACTGTTTGGCCGAGAG	2780
451	IleValAlaIleAspAlaLeuHisPheAlaGlnSerHisHisGlnTyrArgGluAspLeu	470
2781	ATCAGACGGGAGCTTAAAGAGCTTACTGTGGATTTCTTCGT-----	2822
471	MetGluArgGluLeuAsnLysAlaTyrIleGlyPheValHisTyrMetValThrProPro	490
2823	CCTGGAGTTTCTTCAGAGAACCTGCTGCAGTGGCTACAGAACTCGGGCTGTGGTGCC	2882

491 ProGly-----ValAlaThrGlyAsnTrpGlyCysGlyAla 502  
2883 TTGGGGGTGATCTAGACTAAAGCTTAAATACAGATCTCTGCAGCTGCTGTAGCTGAG 2942  
503 PheGlyGlyAspSerTyrLeuLysAlaLeuLeuGlnLeuMetValCysAlaGlnLeuGly 522  
2943 CGAGAGCTGCTTATTTTACCTTTGGGACTCAGAACTGATGAGACACATTTACAGCATG 3002  
523 ArgProLeuAlaTyrThrPheGlyAsnValGluPheArgAspPheHisGluMet 542  
3003 CATACATCTCCTCAGAGGAACTGACTGTGGAGAGTATATAGCTGTGCTTACCA 3062  
543 TrpLeuLeuPheArgAsnAspGlyThrValGlnGlnLeuTrpSer---lleLeuArg 561  
3063 TATTACAATGAAGATGCAAGAACTCTCCACCCCGGACCA----- 3104  
562 SerTyrSerArgLeuLeuLysGlyLysSerSerLysGluProArgGluAsnLysAlaSer 581  
3105 GACATCAAGCTTTATCCATTCATATACCATGAGTTGAGTCTGTACACAGACCACCAAC 3164  
582 LysLysLysLeuTyrAspPhele-----LysGluLysLysValArgAsp 598  
3165 CAGCCGGGCAAGAGGACGGGGGCC 3188  
599 ValProGlyGluGlyAlaSerAla 606

## RESULT 13

US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-302-812-10

Alignment Scores:  
Pred. No.: 1.37e-35 Length: 726  
Score: 465.00 Matches: 204  
Percent Similarity: 38.19% Conservative: 116  
Best Local Similarity: 24.34% Mismatches: 282  
Query Match: 6.42% Indels: 236  
DB: 4 Gaps: 35

US-09-302-812-1 (1-4070) x US-09-302-812-10 (1-726)

852 GACCAAAATAGTGAGGAGAGTAGATAATACAGCAGTTTTCACACATGTAAAGCTTGGC 911  
10 AspProValThrGlnAspGluLysAspTyrGluAspTyrVal---GlyValGlyPheAla 28  
912 AATGCAAGCAGACGATGAA-----GATGAACAGGGGAGAGAGCCAGCAGC 959  
29 HisGlnValProThrMetLysArgArgLysLeuThrGluHisGlyAsnThrGluSer 48  
960 CACAGAGAGTGGCAGGCTTGCCATCTGCAGAAAGCCTGTGAGGGGTGCAGAGGAG 1019  
49 LysGlu-----AspProGluGlu-----ProLys 56

1020 GAGACAGAGCTGGTGTCCGAGAGCCCTTGTCCGACACACTGGCTCTGAGGAT----- 1070  
57 SerArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGlu 76  
1071 -----CTTGGTACTGGACTGAAATAATGCC----- 1094  
77 AsnProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLys 96  
1095 -----AACGATTGCAATAGACAAAGAAAGTAGTCTAGCAAAAT 1130  
97 IleSerAsnIleGluSerLeuAspAsnValThrGluArgSerGluHisThrLeuAspAsn 116  
1131 -----TCTCTCCATTTGAGAAAGAAAGTGAACCTGAGTCCCAATGGATGTA----- 1178  
117 HisLysSerThrGluProMetGluLysValAsnAsnLysSerAsnIleAspValAla 136  
1179 -----GATAATTCAAAATAATGTGTGAGGAT 1205  
137 IleAsnSerAspGluAspGluLeuValLeuGluGluAsnAsnLysGluMetArgAsp 156  
1206 TCAGAGAGCAGATGAAGACAGACAGTCCAGGTTTT-----GATGAACAGGAA----- 1250  
157 GlyGluGlnValGlnGlnLeuSerGlnAspLeuPheAlaAspAspGlnGluLeuGlu 176  
1251 -----GATAGCAGTTCTGCTCAA 1268  
177 TyrProGlyIleMetLysAspThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
1269 ACAGCAATAAACCCTTCAAGGTTCCAAAGAGAGAGTCCAGCTGAGTGTGAGGAAGCGG 1328  
197 ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
1329 TCCTCTGCTAAGGGAGGTGAGATTGCGATTACATTTCCAAATTTGAAGAGGAGAGAGTCA 1388  
212 -----ThrPheValGlyGluAspSerLys 219  
1389 CTTGGAATGAATGATGTGAATGCCAAGAGCTGGAAGTACTTCTAGCTGTAATAGTAG 1448  
220 Ala-----ThrLysThrValArgThrSerSerSerPheLeuSerThr--- 234  
1449 TGCAGAAATTTCTAAGCAACATGGGAGAGAGGATCTTCAAAATACACAGATCATTTCTATGAGA 1508  
234 ----- 234  
1509 GTGCCCCAAGCAGAGCAGCAAGAAAGAAACAATGTGAATGAAACATCAAGACAGAGAA 1568  
235 ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
1569 AGGAAGATCCCTAATAATACATTCACCTCACCCTTCT-----CCAGATAAGAAATGGCTT 1622  
255 LysHisValIleAlaPheThrGluGluAsnLeuThrLeuGlnProAsp----- 270  
1623 GGAATCTCTATTGAGGAGATGAGGAGAGATGCCAAGGTGTGGGATCCGGCTCCCTCCCTG 1682  
271 -----LeuAsnLysVal 274  
1683 AGACCATCTCCATCAATCAGAGTACTATTGCGGTAGATCTTTTCCGAATAGGAGAGTT 1742  
275 AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
1743 CCTAAACCTTTCCCA---ACACATTTTAAAGATTGTGGGACACACAGCAT----- 1790  
286 ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304  
1791 GTTAAGATCGCTTGTTCAGAACAAACTTGTACCTCTGGGAAGATGAGATGGTGAGCGA 1850  
305 IleValLeuProGlnArgTyrArgGluPhe-----AspSerArgGlyArgArg 320  
1851 GCTGCAGGCGCCGCTGGGAACATCATTCAGACTGCATTTCTCAACAGGCTCATCGGCC 19-  
320 -----  
1911 CAGAACCTGAAGGATGCTATTCTGAAGTACAAT-----GTGGCATATTCTAAGP

[illegible]

1179 QY -----GATAATTCACAAATAGTTGTCCAGAT 1205  
 137 Db IleAsnSerAspGluAspGluValLeuGluGluAsnLysGluMetArgAsp 156  
 1206 QY TCAGAAGCAGATGAAGACAGACAGTCCAGGTTTT-----GATGAACAGGAA----- 1250  
 157 Db GlyGluGlnValGlnLeuSerGlnAspLeuPheAlaAspAspGlnLeuLeuGlu 176  
 1251 QY -----GATAGCAGTTCTGCTCAA 1268  
 177 Db TyrProGlyIleMetLysAspThrThrGlnLeuAspIleThrAspSerGluValGlu 196  
 1269 QY ACAGCAATTAACCTTCAGGTTCCAAACCAAGAGAGCTGCACCTGAGTTGAGGAAGCGG 1328  
 197 Db ThrAlaGlnLysMetGluMetIleGluGluThrGluAlaAspSer----- 211  
 1329 QY TCCTCTGCTAAGGAGGTGAGATTCGATTACATTTCCAAATTTGAAGAGGAGAGAGTCGA 1388  
 212 Db -----ThrPheValGlyGluAspSerLys 219  
 1389 QY GCTGAATGAATGATGTGAATGCCAAACGACCTGGAAGTACTTCTAGCTGATGTAGAG 1448  
 220 Db -----ThrLysThrValArgThrSerSerSerPheLeuSerThr--- 234  
 1449 QY TGCAGAAATTTCTAAGCAACATCGGAGAAAGGATTTCTAAATCACAGATCATTTTCATGAGA 1508  
 234 Db ----- 234  
 1509 QY GTGCCCAAGACAGAGACAAAAGAAAGCAATGTGAATGAACATCAAGAGACAGAA 1568  
 235 Db ValSerThrCysGluAlaProAlaLysGlyArgAlaArgMetTyrGlnLysGluLeuGlu 254  
 1569 QY AGGAATCCCTTAATACATATTCACCTCACCTTCT-----CCAGATAAGAAATGGCTT 1622  
 255 Db LysHisValIleAlaPheThrGluGlyAsnLeuThrLeuGlnProAsp----- 270  
 1623 QY GGAACCTCTATTGAGGAGATGAGAGAAATGCCAAGGTGGGATCGGCTGCCCTCCCTTG 1682  
 271 Db -----LeuAsnLysVal 274  
 1683 QY AGACATCTGCCAATCACACAGTACTATTTCGGGTAGATCTTTTGGAAATAGAGAAGTT 1742  
 275 Db AspProAspArgAsnTyr-----ArgTyrCysThrIle 285  
 1743 QY CCTAACCTTCCCA---ACACATTTTAAAGATTTGTGGGCAACAAGCAT----- 1790  
 286 Db ProAsn---PheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGlyProLys 304  
 1791 QY GTTAAGATGCCCTGTTCAGAACAAACTTGTACCTGTGGAAAGATGAGAATGGTGAGCGA 1850  
 305 Db IleValLeuProGlnArgTrpArgGluPhe-----AspSerArgGlyArgArg 320  
 1851 QY GTCGAGGAGCGCGGTGGGAACCTCATTCAGATGCACCTCTCAACMGCTCACTCGGCC 1910  
 320 Db ----- 320  
 1911 QY CAGAACCTGAAGCTCTATTCTGAAGTACAAT-----GTGGCATATTTCAAGAA 1961  
 321 Db -----ArgAspSerTyrPheTyrPheLysArgLysLeuAspGlyTyrLeuLysCys 337  
 1962 QY TGGGACTTTTACAGCT-----TTGATTGATTCTTGGAT----- 1994  
 338 Db TyrLysThrThrGlyTyrPheMetPheValGlyLeuLeuHisAsnMetTrpGluPheAsp 357  
 1995 QY ---AAGGTACTGAAGACAGAGCTCAACCTGTATCATCGCTCCATCTTGCCTGATATG 2051  
 358 Db ProAspIleThrTyrLysLeuProAlaLeuGluMetTyrTyrLysGluMetSerGluLeu 377  
 2052 QY GTG-----AAAATTCGACTCTCTCTG 2072  
 378 Db ValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAlaLysThrAla 397  
 2073 QY CCAAAATTTGTACCCAGCCCAATFACCCTCTCTGAAACAGAGATGAATCATTCATCCATCACA 2132

398 Db GluAspIleLeuProGluArgIleTyrArgLeuValGlyAspVal---GluSerAlaThr 416  
 2133 QY ATGTCACAGGAACAGATTGCCAGCTTTTACGTAATGCTTTCTCTGACGTTTCCAGCA 2192  
 417 Db LeuSerHisLysGlnCysAlaLeuValAlaArgMetPhePhe----- 431  
 2193 QY CGCAATGCCAAGATGAAATCATCAGATTATCCAGTTATCCAGATATTAATCTCAATCGGTTG 2252  
 432 Db -----AlaArgProAspSerProPheSer-----PheCysArgIle 443  
 2253 QY TTTGAGGACGCTTCATCAAGGAACACAGAGCTTAAACGCTTCTCTGCTACTTTAGA 2312  
 444 Db LeuSerSerAspLysSerIleCysValGluLysLeuLysPheLeuPheThrTyrPheAsp 463  
 2313 QY AGAGTCACAGAGAAAAACCCACTGGGTGTGTGACATTC-----ACAAGACAGAGTCTT 2366  
 464 Db LysMetSerMetAspProProAspGlyAlaValSerPheArgLeuThrLysMetAspLys 483  
 2367 QY GAAGATTTTCAGAGTGGGAAGATGTCAAAAACCTCTCGACTCGACTGCATGTCTACTTAC 2426  
 484 Db AspThrPheAsnGluGluTrpLysAspLysLysLeuArgSerLeuProGluValGluPhe 503  
 2427 QY -----GAAGGTACCATAGAAGAAACCGCCAGGCGCATGCTACAGGTGGATTTTCAAC 2480  
 504 Db PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAsn 522  
 2481 QY CGTTTCGTTGGAGTGGTGTAAACACAGTCGACGACTTGTCAAGAGAAATCCGCTTTTATA 2540  
 523 Db GluHisLeuGlyGlyGlyValLeuAsnHisGlySerValGlnGluGluLeuLeuPheLeu 542  
 2541 QY ATCAACCTCGATTGATTGTTTTCACGGCTCTTCACTGAGGTGCTGCTGATCACAATCAATGT 2600  
 543 Db MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla 562  
 2601 QY CTTATCATCAGAGTACTGAGCAGTACAGTACAGTACAGAGCTATGCCAAACATACCGC 2660  
 563 Db IleSerIleValGlyAlaTyrValPheSerSerTyrThrGlyTyrGlyHisThrLeuLys 582  
 2661 QY TGGGCC-----CGGAGCCATGAAGACAGAGCGAA---AGGACGACCTGG 2702  
 583 Db TrpAlaGluLeuGlnProAsnHisSerArgGlnAsnThrAsnGluPheArgAspArgPhe 602  
 2703 QY CAGAGCGCAGCAGCTAGATGCTGCCATCGACGCCCTCCACTTCAGACGC----- 2753  
 603 Db GlyArgLeuArgValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu 622  
 2754 QY -----TACCTCGACAGTTGTGCCCGAGAGATCAGACGGGAGCTTAACAGAGCTTAC 2807  
 623 Db AspCysGlnThrGluGlnLeuAsnLysAlaAsnIleIleArgGluMetLysLysAlaSer 642  
 2808 QY TGTGATTCTTCGTCCTGGAGTTCTTTCAGAGAACCTGTCTGCGAGTGGCTACAGAAAC 2867  
 643 Db IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp 661  
 2868 QY TGGGCTGTGGTCCCTTTTGGGGTGATGCTAGCTTAAAGCCTTAATCAGATCTCTGCA 2927  
 662 Db TrpGlyCysGlyAlaPheAsnGlyAspLysProLeuLysPheIleIleGlnValIleAla 681  
 2928 QY GCTGCTAGCTGACGAGCAGCTGTTTATTTTCACCTTTGGGAGCTCAGAACTGATGAGA 2987  
 682 Db AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla 701  
 2988 QY GACATTTACAGCATCATACATCTCTCTGAGAGAGAAACCTGACTGTTGGAGAA 3041  
 702 Db LysCysLysLysIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys 719

RESULT 15

US-09-511-507-10

; Sequence 10, Application US/09511507

; Patent No. 6395543

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.



Db	464	LyseMetSerMetAaspProAaspGlyAlaValSerPheArgLeuThrIysMetAspIys	483
QY	2367	GAAGATTTCCAGAGTGGAAAGAAATGCTGAAAAATCTCTGACTCGATCGATGCTCACTTAC	2426
Db	484	AspThrPheAanGluIutPlyAspLysLeuAargSerLeuProGluValGluPhe	503
QY	2427	-----GAAGGTACCATAGAAAGAAACGGCCAGGCATGCTACAGTGGATTTCGAAC	2480
Db	504	PheAspGluMetLeuIleGluAspThrAlaLeu---CysThrGlnValAspPheAlaAan	522
QY	2481	CGTTTCGTGGAGTGGTAAACCACTGCAGGAGCTTGTGCAAGAAAGAAATCCGCTTTT	2540
Db	523	GluHisLeuGlyGlyValLeuAanHisGlySerValGlnGluGluIleargPheLeu	542
QY	2541	ATCAACCTCAGTTGATTGTTTCACGGCTCTTCAGTGGTGTCTGGATCAACATGAATGT	2600
Db	543	MetCysProGluMetMetValGlyMetLeuLeuCysGluLysMetLysGlnLeuGluAla	562
QY	2601	CTTATCATCACAGTACTGACGACGTACAGTGAATACACAGGCTATCCGAAACATACCCG	2660
Db	563	IleSerIleValGlyAlaItyrValPheSerSerItyrThrGlyTyrGlyHisThrLeuLys	582
QY	2661	TGGGCC-----CGGAGCCATGAAGACAGGAGCCAA--AGGACGACACTGG	2702
Db	583	TrpAlaGluLeuGlnProAanHisSerArgGlnAsnThrAandGluPheargAspArgPhe	602
QY	2703	CAGAGCGGCACGACTGAGATCGTGGCCATCGACGCCCTCCACTTCAGACGC-----	2753
Db	603	GlyArgLeuargValGluThrIleAlaIleAspAlaIleLeuPheLysGlySerLysLeu	622
QY	2754	-----TACCTGCACCAATTTGTGGCCCGCAGAGATCAGACGGGACTTAACAGGCTTAC	2807
Db	623	AspCysGlnThrGluGlnLeuAanLysAlaAanIleIleargGluMetLysLysAlaSer	642
QY	2808	TGTGGATTTCTCGTCTCTGGAGTTTCTTCAGAGAACCTGTCTGAGTGGCTACAGAAAC	2867
Db	643	IleGlyPheMetSerGlnGlyProLysPheThrAsnIle---ProIleValThrGlyTrp	661
QY	2868	TGGGGCTGTGCTTGTGGGGGTGATGTAGACTAAAGCCCTTAATACAGATCCCTGGCA	2927
Db	662	TrpGlyCysGlyAlaPheAanGlyAspLysProLeuLysPheIleIleGlnValIleAla	681
QY	2928	GCTGCTGTAGTCAGCGACGCTGGTTTATTTCACCTTGGGGACTCAGAACTGATGAGA	2987
Db	682	AlaGlyValAlaAspArgProLeuHisPheCysSerPheGlyGluProGluLeuAlaAla	701
QY	2988	GACATTTACGATCATCATCTCTCAGTGAAGGAACTGACTGTTGGAGAA	3041
Db	702	LysCysValLysValIleIleGluArgMetLysGlnLysAspValThrLeuGlyLys	719



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 15:32:43 ; Search time 34.0417 Seconds  
(without alignments)  
12447.891 Million cell updates/sec

Title: US-09-302-812-3  
Perfect score: 7208  
Sequence: 1 ggcgtctggagtgaggag.....agaaaaaaaaaaaaa 4069

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p model -DEV=xlp  
-Q=/cgn2 1/USPTO spool p/6333148/runat 26052004 150053 5664/app query.fasta 1.12437  
-DB=SwissProt\_42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -iLOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=6333148 -CGN 1 1 128 @runat 26052004 150053 5664 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	186.5	2.6	2476	1 ATRX_MOUSE	Q61687 mus musculus
2	181.5	2.5	1359	1 ATRX_CAEEL	Q9u7e0 caenorhabdi
3	181.5	2.5	5085	1 PCLO_RAT	Q9jks6 rattus norv
4	174	2.4	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
5	172.5	2.4	5038	1 PCLO_MOUSE	Q9cyx7 mus musculus
6	162	2.2	2492	1 ATRX_PANTR	Q7yqm4 pan troglod
7	161	2.2	2492	1 ATRX_HUMAN	P46100 homo sapien
8	161	2.2	2492	1 ATRX_PONPY	Q7yqm3 pongo pygma
9	160	2.2	1957	1 SPOF_SCHPO	Q10411 schizosacch
10	159.5	2.2	1183	1 BMS1_YEAST	Q08965 saccharomyc
11	159	2.2	406	1 YG9_SCHPO	P32583 schizosacch
12	159	2.2	1031	1 YG9_SCHPO	P10496 schizosacch
13	159	2.2	1462	1 NKCR_HUMAN	P30414 homo sapien
14	159	2.2	2418	1 SPCA_HUMAN	P02549 homo sapien
15	158.5	2.2	6632	1 UN89_CAEEL	O01761 caenorhabdi
16	158	2.2	1311	1 ATRX_DROME	Q99qn5 drosophila
17	155	2.2	1138	1 BM2K_MOUSE	Q91z96 mus musculus
18	154.5	2.1	646	1 SGI_BOVIN	P23389 bos taurus

19	154.5	2.1	3259	1 GOB1_HUMAN	Q14789 homo sapien
20	154	2.1	1972	1 MYHB_MOUSE	Q08638 mus musculus
21	154	2.1	2238	1 GOA4_MOUSE	Q91vw5 mus musculus
22	153.5	2.1	1878	1 BRC1_CANFA	Q95153 canis famil
23	152.5	2.1	2415	1 SPCA_MOUSE	P08032 mus musculus
24	152	2.1	644	1 NFM_RABIT	P54938 oryctolagus
25	152	2.1	1447	1 BUD4_YEAST	P47136 saccharomyc
26	152	2.1	5560	1 SPEN_DROME	O8sx83 drosophila
27	151.5	2.1	1338	1 ACIN_MOUSE	Q9jix8 mus musculus
28	150.5	2.1	1312	1 RASO_YEAST	P12753 saccharomyc
29	150.5	2.1	1972	1 MYHB_RABIT	P35748 oryctolagus
30	150	2.1	2137	1 SPCB_HUMAN	P11277 homo sapien
31	150	2.1	3969	1 HRX_HUMAN	Q03164 homo sapien
32	149.5	2.1	1210	1 AF4_HUMAN	P51825 homo sapien
33	149	2.1	1093	1 KYD5_CAEEL	Q03563 caenorhabdi
34	148.5	2.1	472	1 TWIE_CAEEL	Q23525 caenorhabdi
35	148.5	2.1	677	1 SGI_HUMAN	P05060 homo sapien
36	148.5	2.1	1928	1 MYS1_YEAST	P08964 saccharomyc
37	148.5	2.1	1969	1 MYS4_CAEEL	P12844 caenorhabdi
38	148	2.1	627	1 ABPX_YEAST	Q08641 saccharomyc
39	148	2.1	1679	1 GCC2_MOUSE	Q8chq3 mus musculus
40	148	2.1	3210	1 CENF_HUMAN	P49454 homo sapien
41	147	2.0	943	1 YMI1_YEAST	P39523 saccharomyc
42	146.5	2.0	667	1 CYLI_BOVIN	P35662 bos taurus
43	146.5	2.0	1560	1 TENN_MOUSE	Q80271 mus musculus
44	146.5	2.0	3214	1 BPAL_HUMAN	Q03001 homo sapien
45	146	2.0	5120	1 PCLO_CHICK	Q9pu36 gallus gall

ALIGNMENTS

RESULT 1  
ATRX\_MOUSE  
ID ATRX\_MOUSE STANDARD; PRT; 2476 AA.  
AC Q61687;  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Transcriptional regulator ATRX (X-linked nuclear protein)  
DE (heterochromatin protein 2) (HPI alpha-interacting protein) (HP1-BP38 protein).  
DE GN ATRX OR XNP OR HP1BP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98213653; PubMed=9545503;  
RA Picketts D.J., Tastan A.O., Higgs D.R., Gibbons R.J.;  
RT "Comparison of the human and murine ATRX gene identifies highly conserved, functionally important domains";  
RL Mamm. Genome 9:400-403(1998).  
RN [2]  
RP SEQUENCE OF 325-1176 FROM N.A.  
RX MEDLINE=97133299; PubMed=8978696;  
RA le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H.,  
RT Jeanmougin F., Lissou R., Chambon P.;  
RT "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors.";  
RL EMBL J. 15:6701-6715(1996).  
RN [3]  
RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC HETEROCHROMATIN.  
RX MEDLINE=20040663; PubMed=10570185;  
RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,  
RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
RT "Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
RL CC -!- FUNCTION: Could be a global transcriptional regulator. Modifies

CC gene expression by affecting chromatin.  
CC -|- SUBUNIT: Probably binds E2H2. Binds annexin V in a calcium and  
CC phosphatidylcholine/phosphatidylserine-dependent manner (By  
CC similarity).  
CC -|- SUBCELLULAR LOCATION: Nuclear. Associated with pericentromeric  
CC heterochromatin during interphase and mitosis, probably by  
CC interacting with Hpl.  
CC -|- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.  
CC -|- SIMILARITY: Contains 1 PHD-type zinc finger.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF026032; AAC08741.1; -;  
CC DR EMBL; X99643; CAA67962.1; -;  
CC DR MGD; MGI:103067; Atrx.  
CC DR GO; GO:000228; C:nuclear chromosome; IDA.  
CC DR InterPro; IPR001410; DEAD.  
CC DR InterPro; IPR001650; Helicase\_C.  
CC DR InterPro; IPR00330; SNF2\_N.  
CC DR Pfam; PF00271; helicase\_C; 1.  
CC DR Pfam; PF00176; SNF2\_N; 1.  
CC DR SMART; SM00487; DEXDC; 1.  
CC KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;  
CC Zinc-finger.  
CC FT ZN FING 219 267 PHD-TYPE.  
CC FT NP BIND 1579 1586 ATP (POTENTIAL).  
CC FT SITE 1704 1707 DEGH BOX.  
CC FT DOMAIN 319 322 POLY-SER.  
CC FT DOMAIN 735 738 POLY-SER.  
CC FT DOMAIN 1001 1004 POLY-GLU.  
CC FT DOMAIN 1130 1135 POLY-SER.  
CC FT DOMAIN 1182 1185 POLY-SER.  
CC FT DOMAIN 1238 1245 POLY-ASP.  
CC FT DOMAIN 1484 1487 POLY-GLU.  
CC FT DOMAIN 1924 1931 POLY-SER.  
CC FT DOMAIN 2205 2208 POLY-LYS.  
CC FT DOMAIN 2245 2248 POLY-GLU.  
CC FT DOMAIN 2403 2408 POLY-GLN.  
CC SQ SEQUENCE 2476 AA; 278601 MW; 90A42B790FC4FF4C CRC64;  
  

Alignment Scores:  
Pred. No.: 0.000113 Length: 2476  
Score: 186.50 Matches: 216  
Percent Similarity: 35.48% Conservative: 142  
Best Local Similarity: 21.41% Mismatches: 379  
Query Match: 2.59% Indels: 273  
DB: 1 Gaps: 53

US-09-302-812-3 (1-4069) x ATRX\_MOUSE (1-2476)

QY 67 AGGCGGCGGTGAGAGAGAGCGCGAGTCCATCTCTCTCAGGTAGTGAATGAGGCTCTC 126  
Db 967 ArgProGlyThrLysGlyLysAlaProAspLeuLysGly-----GluThrLeu 983  
QY 127 CGC-----GGGGCGGCGCGGAGCATGCGCTGCTGCTGCCAGCAT 168  
Db 984 LysArgGluGlnIuTrpAspSerSerAspGlyThrGluArgLeuProGluGlu 1003  
QY 169 GAATGCGGCGCGCGGTGTGAACCTCCACCAAGC-----GAC 207  
Db 1004 GluIleGlyProPheSerLysGlyLysGlnSerLysThrAspThrAlaGlyGlyGlu 1023  
QY 208 CCGCTGGGCGCGCGCTACAACTCCGCGGTGCTTCGGACGCCCGCGAGCTTCGAGCAG 267  
Db 1024 LysLysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1043  
QY 268 GCAGAGGCGCGCTCTCGAGCCCAAGGAGCGCTCAGTCGAGTTGAGGTCGCCAGCTCTC 327

Db 1044 ValAspLysLeuPro---GlyLysGlyAspSer---CysAspSerSerGluAspLysLys 1062  
QY 328 GCCAG-----CCTGCG---TCCCAGG 345  
Db 1062 hrArgAsnArgValSerLeuArgGluLysLysArgPheSerLeuProAlaLysSerPro 1082  
QY 346 GCAGGCGG-----G 354  
Db 1082 LysArgProGluCysSerSerSerAspThrGluLysSerLeuLysGlyGlnCys 1102  
QY 355 ACAGCAGAGCGGCGCCACCTCGCTGTGTT---TTCACAAACAAAGACTATTACCATT 410  
Db 1102 spSerThrGluLysArgProLysArgLysLeuArgGluArgAsnSerSerSerL 1122  
QY 411 GGATGGACATAAGGAATCAAGACAGCGGAATCAGAAAGT---TTGGATAGTAAG 464  
Db 1122 ysArgAsnThrLysGluValLysSerAlaSerSerSerSerAlaGluGlySerSer 1142  
QY 465 AAAACAACAATACAGAAATCAATCCATGATGATCTCTGTACAAAAGATACTTTTACC 524  
Db 1142 LuAspAsnLysGlyGlnLysLysGlnArgThrSerAlaLysLysLysThrGlyAsnThr 1162  
QY 525 AACATAATGTAGAAAAATAGTAAATGTTTCTCAGCTAAGTCTTGATAGTCACTCACTG 584  
Db 1162 ysGluLysLysArgAsnSerLeuArgAlaThrProLysArgLysGlnValAspLysThr 1182  
QY 585 AAAAAGTACACAGTATTGTAACAGCATCATGAGCAATGTGTAAAGTGGCAAAATG 644  
Db 1182 erSerSerSerAspLysAspAspGlnAsnSerAlaGluGluSerSerAspG 1202  
QY 645 AAGGGA---CACAG-----CACAG----- 658  
Db 1202 LuGlnLysLysProValThrGluAsnLeuValLeuProSerHisThrGlyPheCysG 1222  
QY 659 -----GAGCAGCTTTTGGAAAGTGAACCTCAACAGTAACTCCCTGGTACCAG 704  
Db 1222 InSerSerGlyAspGluAlaLeuSerLysSerValProAlaThrVal----- 1237  
QY 705 AGCAGTTTGTATGCTAATCGTACATTCGTCACCTCAAAATGATGATCAGTACAGTACAG 764  
Db 1238 -----AspAspAspAspAspAsp 1244  
QY 765 ATAGTGAAGAGATAGAGACATCAACAGTCTTCTCAACTGTAAAGTTGCAATGCAA 824  
Db 1244 snAspProGluAsnArgLysLysMetLeuGluGluLysAlaAsnLeu 1264  
QY 825 AGCAG-----ACTACGGAAGTGAACACGCCAGAGAGCCAAAGCCACCAG 872  
Db 1264 erSerAspGluAspGlySerSerAspAspGluProAspGlyGlyLysLysArgLleG 1284  
QY 873 AGTGCAGCAAGTCTTCCCATCTCGGGAAGAGCTGTGCAAGTTGTCTCAGCAAGATGAGATAG 932  
Db 1284 LysLysGlnSerGluGluSerProAlaAspAsp---GlyGluLeuArgArgGluGlnLeu 1303  
QY 933 AGCTGTGTCGAAGAGTCCATTGTCTCAGATGTTGGCTCTCAGGATGTTGGTACTGGTCAA 992  
Db 1303 laValAsnGlnValAsnSerGluSerAspSerAspSerGluGlu-----SerL 1319  
QY 993 AAAATGAC-----AACAAATTGATTAGACAGAAAGTTGCCCTAGCAAAATCTCTCTC 1043  
Db 1319 ysLysProArgTyrArgHisArgLeuLeuArgHisLysLeuThrLeuSerAspGlyGln 1339  
QY 1044 CA---TTTCAGAGGAAGTGAACCGAATCCCGATGATGGGATAATCTTAAATAA 1100  
Db 1339 erGlyGluGluLysProThrLysProLysGluHisLysGluAlaLysGlyArgAsnArg 1359  
QY 1101 GTTGTCAAGACTCAGAACGAGATGAGGAGACAGTCCAGTCCAGTTTGTATGAACAGAGATG 1160  
Db 1359 rGlyValSerSerGluAspSerGluAspThr-----AspPheGlnGluSerG 1375  
QY 1161 GT-----AGTTCCTCCCAACACAGCAATAAACCCTTCAAGTTCACAGCAA 1205

QY	2178	GGAAACCGGAGAAACTTAAACCGCTCTTCTGCTACTTCTTAAAGAGAGTCACAGAGAAAAAAC	2233
Db	1681	--LysSerArgLysLeuLysAspIlePhe-----AsnLysAlaLeuValAspProGlyP	1698
QY	2238	CTACTGGTGTGGTGACATTTACAAGACACAGAGTCTTGAACAGATTTTCCAGAAATGGGAAAGAT	2297
Db	1698	roAspPheValValCysAspGluGlyHisIleLeuLysAsn-----GluAlaSerAlaVal	1716
QY	2298	GTGAAAAACCCCTTGG-----ACAGATTGCATGTCTACTTACAGAGGTGACC----	2341
Db	1716	alSerLysAlaMetAsnSerIleLysSerArgArgIleLeuLeuThrGlyThrProL	1736
QY	2342	-----ATAGAAGAAAAATGGCCCAAG	2360
Db	1736	euGlnAsnAsnLeuLeuGluTyHisCysMetValAsnPheIleLysGluAsnLeuLeuG	1756
QY	2361	GCATGTCTACAGGTGGATTTTGCAAATCGTTTTTCTT-----GGAGGTGGTGTTAA	2408
Db	1756	lySerIleLys---GluPheArgAsnArgPheIleAsnProIleGlnAsnGlyGlnCysA	1775
QY	2409	CCAGTGCAGGACTTGTGCAAGAGAAATCCGCTTTTAACTCAATCTCGACTTGATTTATTT	2468
Db	1775	laAspSerThrMetVal-----AspValArgValMetLysLysArgAlaHisIleLeu	1793
QY	2469	CACGGCTC-----TTCACTTGAGGTGCTG--	2491
Db	1793	yrGluMetLeuAlaGlyCysValGlnArgLysAspTyrThrAlaLeuThrLysPheLeuP	1813
QY	2492	-----GATCAATCAATGCTCTAATTATCACAGGTACTGAG-----CACTACAGTGAATACA	2543
Db	1813	roProLysHisGluTyrValLeuAlaValArgMetThrAlaIleGlnCysLysLeuTyrG	1833
QY	2544	CAGGCTATGTCTGACATATCGTTGGTCCCGGACCCACAGAGATGCGAGTCAAGGGGACG	2603
Db	1833	lnTyrTyr-LeuAspHisLeuThrGlyValGlyAsnSerThrGluGlyGlyArgGly---	1851
QY	2604	ACTGGCAGCGCGCTGCACTG	2624
Db	1852	---IysAlaGlyAlaLysLeu	1857
RESULT 2			
ATRX_CABEL			
ID	ATRX_CABEL	STANDARD;	PRT; 1359 AA.
AC	Q9U7E0; Q02061;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).		
GN	XNP-1 OR B0041.7.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N. A., AND CHARACTERIZATION.		
RX	MEDLINE=99365296; PubMed=10433961;		
RA	Villard L., Fontes M., Ewbank J.J.,		
RT	"Characterization of xnp-1, a Caenorhabditis elegans gene similar to		
RT	the human XNP/ATR-X gene."		
RL	Gene 236:13-19(1999).		
RN	[2]		
RP	SEQUENCE FROM N. A.		
RC	STRAIN=Bristol N2;		
RA	Fulton R., Woldmann P.;		
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.		
CC	-I- FUNCTION: Could be a global transcriptional regulator. Modifies		
CC	gene expression by affecting chromatin (potential).		
CC	-I- SUBCELLULAR LOCATION: Nuclear.		
CC	-I- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.		

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
DR EMBL; AF134186; AAD55361.1; -;  
DR EMBL; AF000196; RAC24256.1; -;  
DR PIR; T34036; T34036  
DR WormPep; B0041.7; CE17314.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELIC; 1.  
DR PROSITE; PS00690; DEAH ATP HELICASE; FALSE NEG.  
KW DNA repair; Hydrolase; Helicase; Nuclear protein; ATP-binding;  
KW DNA-binding.  
FT NP\_BIND 496 503 ATP (POTENTIAL).  
FT SITE 636 639 DEAH BOX.  
FT FT 67 70 POLY-ASP.  
FT DOMAIN 266 272 POLY-GLU.  
FT DOMAIN 276 281 POLY-LYS.  
FT DOMAIN 372 375 POLY-LYS.  
FT DOMAIN 603 608 POLY-LYS.  
FT DOMAIN 859 862 POLY-LYS.  
FT CONFLICT 479 479 C -> F (IN REF. 2).  
FT SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;  
SQ

Alignment Scores:  
Pred. No.: 0.000209 Length: 1359  
Score: 181.50 Matches: 182  
Percent Similarity: 32.44% Conservative: 131  
Best Local Similarity: 18.86% Mismatches: 328  
Query Match: 2.52% Indels: 324  
DB: 1 Gaps: 39

US-09-302-812-3 (1-4069) x ATRX\_CABEL (1-1359)

QY 78 GAGAGAGCGCGAGTCACTCTCTCAGGTAGTGAATGAGGCTCTCCGGGCGCGG 137  
DB 33 GluLysArgAlaGlnLysLeuLysGluLysArgGluArgGluLysPro----- 49  
QY 138 CCCGGGACAGTCCGCTGCTGCTCCAGCATGAATGAGCGGCGCCGCTGTGAACCTTCA 197  
DB 50 -----ProPro 51  
QY 198 CCAAGACCGCTGCGGCGCGCTTACAACTTCGCGCGCTGCTCGGAGCGCGAGCT 257  
DB 52 LysLysArgProAla----- 56  
QY 258 TTCGAGGAGGAGGCGCGCTCTCGACCCCAAGACGCTCAGCTGCGAGTTCAGGGTCC 317  
DB 56 ----- 56  
QY 318 CACCGTCTCGCAGCTGCTGCCAGGCGCGGCGGAGCAGCAGAGCGCGCCACT 377  
DB 57 ---LysLysArgLysAlaSerSerSerSerSerSerSerSerSerSerSerSerSerPro 75  
QY 378 CGCTTGTCTTCAACAAAGACTATTACCGAGTTGGATGGACATTAAGAGATCAACAGAC 437  
DB 76 ArgLysSerSerLysLys-----SerArgLysArgAlaLys--S 88  
QY 438 CGGAATCAGAAAGTTGGATAGTAAGAAAACAAATACAGATAGATCCATGATCA 497  
DB 88 erGluSerGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 103  
QY 498 GTTCTGTACAAAGAGATACTTTTACCAACATATGAGAA-----AAATTAGTAATG 551  
DB 103 ysSerLysSerLysLysValAspGlnLysLysLysLysLysLysLysLysLysLysArgT 123

QY 552 TTTCTCAGCTAAGTCTTGATAGTCACTCACTGAAAGAAAGTACACAGTATTTGAACACG 611  
DB 123 hrThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerL 141  
QY 612 ATCAGACGTGCAAGTGTAGTGGCAAAATGAAGGGAACACACGAGCAGCTTTTGG 671  
DB 141 ysLysLysSerLysLysThrLysLysGlnThrSerSerSerSerSerSerSerSerSer 158  
QY 672 AAAGTGAACCTCAAAACAGTAACCTGTTACACAGCAGCTTTAGTAAATGCTAAATGATG 731  
DB 158 luSerGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 178  
QY 732 GGTCACTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 791  
DB 178 ysArgAlaGluThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 198  
QY 792 AGTTCTCACTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 851  
DB 198 ysGlyLeuLys-----LysLysAlaLysSerSerSerSerSerSerSerSerSerSer 216  
QY 852 GAGAGCCAAAGCCACAGAGTGCAGCAAGTCTTCCATCTCTGGGAGAGCTGTGCAA 911  
DB 216 ysGluValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 236  
QY 912 GTTGTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 971  
DB 236 luAspGluAlaProGluLysLysLysLysLysLysLysLysLysLysLysLysLys 256  
QY 972 AGGATGTTGTTACTGGTCAAAAATGATGATGATGATGATGATGATGATGATGATG 1031  
DB 256 luGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 269  
QY 1032 GAAATCTCTCTCCATTTGAGAGAGAAAGTGAACCC-----GAATCAGGATGG 1079  
DB 270 -----GluLysGluSerSerSerSerSerSerSerSerSerSerSerSerSerSer 284  
QY 1080 ATGTGGAATAATCTAAATAATGTTGTCAAGCTCAAGAGAGAGAGAGAGAGAGAGAG 1138  
DB 284 laValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 304  
QY 1139 -----GGTTTGTGATCAACAAGAAGATG 1160  
DB 304 lnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 324  
QY 1161 GTAGTCTCTCCCAACAGCAATAAATCTCAAGGTTCCAAAGAGAGAGAGAGAGAGAGAG 1220  
DB 324 ysSerGluSerGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 341  
QY 1221 AATTAGAAACGGTACTCTACTAAGGCGGTGAGTTAGATTACATTTCATTTGAAG 1280  
DB 342 -----LysLysGlnLysSerSerSerSerSerSerSerSerSerSerSerSerSer 350  
QY 1281 GAGGAGAGAGTGGCAGTCAATGATGATGATGATGATGATGATGATGATGATGATGAT 1340  
DB 350 lysSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 364  
QY 1341 GCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1394  
DB 364 erLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 384  
QY 1395 TCACAGATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1454  
DB 384 euGlnLysGluThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 404  
QY 1455 CCAACATCAAGA----- 1468  
DB 404 ysLysGlnLysGluPheAsnGlyLysLysLysLysLysLysLysLysLysLysLys 424  
QY 1469 -----ACAGAAAGAGAGATCCCTAAATACCTTCCACCTCCTCCTCCAG 1514  
DB 424 euThrGlyThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 441  
QY 1515 ATAAGAGTGGCTTGGAACTCCCATTCAGGAGATGAGAGATGCTCGCTGCTGGGATCC 1574

```

Db 441 sp-----SerSerThrValAspGluGluSerLysLysPro----- 452
QY 1575 GGTCGCTCTCTTGAGACCATCTGCCAATCACACAGTAACATATTCGGGTAGATCTTTTGC 1634
Db 453 -----ValGluValHisAsnSerLeuValA 461
QY 1635 GAGCAGGAGAGATTCTCAACCTTTTCCAAACACATATAAAGATTTCGTGGGATAACAAGC 1694
Db 461 rg-----IleLeuLysProHisGlnAlaHis----- 469
QY 1695 ATGTTAAATGCTTGTTCAGAACAAATTTGACCCAGTGGAGATGAGAATGGTGAGC 1754
Db 470 -----GlyIleGlnPheMetTyrAspCysAlaCysGluSerLeuAspA 484
QY 1755 GA-----ACTGCGGGAGCGGTGGGAGCTCATTCAGATGCACTTCTCAACAATTTA 1808
Db 484 rgLeuAspThrGluGlySer----- 490
QY 1809 CACGACCCCAAACTTGAAGGATGCTATTCTGAAATACAAATGTGGCATATTCCTAAGAAAT 1868
Db 491 -----GlyGlyIleLeuAlaHisCysMetGlyLeuGlyLysThr- 503
QY 1869 GGGACTTACAGCTTTCATGATGTTCTGGGATTAAGTACTTGAAGACGAGACGTCACAC 1928
Db 504 -----LeuGlnValIleThrPheLeuHisThrValLeu----- 514
QY 1929 ATTTATATCATGCTCATCTTCCTGATATGTTGAAATTCGACTCTCTGCTCCA---ANTA 1985
Db 515 -----MethHisGluLysIleGlyGluLysCysLysArgValLeuValValProLysAsnV 534
QY 1986 TTTGCCCCAGCCAAATACCACCTCTCTGAAACAGAGAGATGAATTCATTCACCAATCTGCG 2045
Db 534 alleleAsnTTPhe-----LysGluPheGlnLysTTPLeuValAspAsnA 550
QY 2046 AGGAACAGATGTCAGCTCTTTAGCTAAATGCTTTCTTCTGACATTTCCACGACGAAATG 2105
Db 550 spGluGluLeuAspThrIleAspValAsn----- 559
QY 2106 CTAAGATGAATCGGAGTATTCTAGTTACCCAGACATTAATTCATCGA-----TTCT 2159
Db 560 -----GluLeuAspSerTyrLysThrIleGluAspArgArgAlaLeuL 575
QY 2160 TTGAGGAGCTTCATCAAGAAACCGGAGAACTTAAACCGCTCTCTCTACTATTAGAA 2219
Db 575 ysAlaTriPHisSerLysThrProSerValMetIleIleGlyTyrAspLeuPheArgI 595
QY 2220 GAGTCACA-----GAGAAAAACCTACTCGGTGGTGATTCATTACAA 2261
Db 595 leLeuThrValGluAspAspProLysLysLysLysProLysAsnArgAsnArgLeuG 615
QY 2262 GACAGAGCTTCAAGATTT-----CCAGAAATGGAAAGAT 2297
Db 615 luLysAlaLysGluAspPheArgLysTyrLeuGlnAsnProGlyProAspMetValValC 635
QY 2298 GTGAA-----AAAC 2306
Db 635 ysAspGluAlaHisLysLeuLysAsnAspSerAlaLeuSerLysCysMetValLysI 655
QY 2307 CTTTCACAGGATGTCATCTACCTACGAGGTACCATAGCAAGAAATGCG----- 2356
Db 655 leLeuThrLysArgIleCysLeuThrGlyThrProLeuGlnAsnAsnLeuMetGluT 675
QY 2357 -----CAAGGATGCTA-----CAGGTGGATTTTGCAA 2384
Db 675 yrHisCysMetValAsnPheValLysProGlyLeuLeuGlyThrLysThrGluPheAlaA 695
QY 2385 ATCGTTTCTGAGGAGTGTGTAACGACGACGACCTTGTGCAAGAA-----G 2432
Db 695 snArgPheVal-----AsnIleIleAsnArgGlyArgThrLysAspAlaSerProLeuG 713
QY 2433 AAATCGGCTTTTAAATCAATCTGAGTGAATTTTCACTCGGCTTCTACTGAGGCTGTG 2492

```

```

Db 713 luValSerPheMetLysArgCysHisValLeuTyrAspHisLeuLysLysCysValA 733
QY 2493 ATCACAAATGATCTCTAATATTATCACAGGTACT----- 2524
Db 733 spArgLysAspTyrArgValLeuThrGluAlaIleProProLysGlnGluTyrValIleA 753
QY 2525 -----GACAGTACAGTGAATACACAGCTATGCTGAGACATATCGTT 2567
Db 753 snValArgGlnThrGluArgGlnCysAlaLeuTyr-AsnAlaPheLeuAsnAspIleVal 772
QY 2568 GGTCCCGGAGCCACGAGATGGAGTCAAAGGACGAGCTCGAGCGCGCTGCACTCAGA 2627
Db 773 GlyAsp-----SerGlyLeuSerLysArg 780
QY 2628 TCCTTGGCCA 2636
Db 781 LeuLeuPro 783

RESULT 3
PCLO RAT
ID PCLO RAT STANDARD; PRT: 5085 AA.
AC Q9JXS6; Q9JLT1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Piccolo protein (Multidomain presynaptic cytomatrix protein).
GN PCLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a presynaptic zinc finger protein structurally related to
RT bassoon.";
RL Neuron 25:203-214 (2000).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 1).
RX Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;
RP VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND
RP ALA-4694.
RX MEDLINE=21181819; PubMed=11285225;
RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;
RA "An unusual C(2)-domain in the active-zone protein piccolo:
RT implications for Ca(2+) regulation of neurotransmitter release.";
RL EMBO J. 20:1605-1619 (2001).
CC -! FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking (By similarity).
CC -! SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By
CC similarity).
CC -! SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic
CC junctions.
CC -! ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JXS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JXS6-2; Sequence=VSP 003930, VSP 003931;
CC -! DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -! SIMILARITY: Contains 2 C2 domains.
CC -! SIMILARITY: Contains 1 PDZ/DHR domain.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```





CC trafficking (By similarity).

CC -|- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By

CC similarity).

CC -|- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of

CC synaptic junctions (By similarity).

CC -|- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Comment=Additional isoforms seem to exist;

CC Name=1;

CC IsoId=Q9Y6V0-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9Y6V0-2; Sequence=VSP\_003923, VSP\_003924, VSP\_003925,

CC VSP\_003926, VSP\_003927;

CC Note=No experimental confirmation available;

CC -|- DOMAIN: C2 domain 1 is involved in binding calcium and

CC phospholipids. Calcium binds with low affinity but with high

CC specificity and induces a large conformational change.

CC -|- SIMILARITY: Contains 2 C2 domains.

CC -|- SIMILARITY: Contains 1 PDZ/DHR domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; Y19188; CAB60727.1; -;

CC EMBL; AC004903; AAD20936.1; -;

CC EMBL; AC004886; AAD21789.1; -;

CC EMBL; AB011131; BAA25485.1; -;

CC EMBL; BC001304; AAH01304.1; -;

CC EMBL; AC004082; AAB97937.1; -;

CC F01; T00634; T00634.

CC HSSP; P04410; 1A25.

CC Gensu; HGNC:13406; PCLO.

CC MIM; 604918; -;

CC GO; GO:0005856; C:cytoskeleton; NAS.

CC GO; GO:0045202; C:synaptic junction; ISS.

CC GO; GO:0005509; F:calcium ion binding; ISS.

CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.

CC GO; GO:0005522; F:profilin binding; ISS.

CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.

CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.

CC InterPro; IPR000008; C2.

CC InterPro; IPR001565; Synaptotagmin.

CC PRINTS; PR00360; C2DOMAIN.

CC PRINTS; PR00399; SYNAPTOTAGMIN.

CC SMART; SM00239; C2; 2.

CC PROSITE; PS00499; C2 DOMAIN 1; 1.

CC PROSITE; PS50004; C2 DOMAIN 2; 2.

CC Calcium/phospholipid-binding; zinc; Metal-binding; Zinc-finger;

CC Repeat; Alternative splicing.

CC NON\_TER 1

CC DOMAIN 400 465

CC 10 X 10 AA TANDEM APPROXIMATE REPEATS OF

CC P-A-K-P-Q-P-Q-P-X.

CC C4-TYPE (POTENTIAL).

CC C4-TYPE (POTENTIAL).

CC POLY-PRO.

CC PDZ.

CC C2 DOMAIN 1.

CC C2 DOMAIN 2.

CC S -> SGNIGIRIVGKEIPGHSBIGAVIAKILPGSAGE

CC QTKLMNEG (in isoform 2).

CC /FTid=VSP\_003923.

CC K -> KPTDGTKVSPITGSIQ (in isoform 2).

CC /FTid=VSP\_003924.

CC G -> GQVMVQNAS (in isoform 2).

CC /FTid=VSP\_003925.

CC TAKHS -> SKRRK (in isoform 2).

CC /FTid=VSP\_003926.

FT VARSPLIC 4762 5147 Missing (in isoform 2).

FT SQ SEQUENCE 5147 AA; 563537 MW; CD5D84990498CD3C CRC64;

Alignment Scores:

Pred. No.: 0.000945 Length: 5147

Score: 174.00 Matches: 140

Percent Similarity: 36.67% Conservative: 120

Best Local Similarity: 19.75% Mismatches: 251

Query Match: 2.41% Indels: 198

DB: 1 Gaps: 32

US-09-302-812-3 (1-4069) x PCLO\_HUMAN (1-5147)

QY 191 CCCTGCACCAAGCGACCGCTGGCGGCCCTACACTTCGCGGCTGCTTCGACGCC 250

DB 1227 ProCysThr-----AlaLysProAspGlnGluLysGluAsp 1239

QY 251 CGGAGCTTTCGAGCAGCAGCGCGCTCTCGACCCCAAGGAGCTCAGTCGAGTTC 310

DB 1240 LysSerAspThrSerSerSerGlnGln-----1249

QY 311 AGGGTCCACCGCTCTCCCGACGCTGCTCCAGCGGCGGACACACAGAGGACG 370

DB 1249 -----ProLysSerProGlnGlyLeuSerAspThrGlyTyrSerSerAspGlyLe 1265

QY 371 GCCACCTCGCTGTTTTCACCAACAAAGACTATTACCACTTGATGGACACT----- 421

DB 1266 SerSerSerLeu-----GlyGluIleProSerLeuIleProThrAspGluLys 1281

QY 422 -----AAAGGAATCAAG-----ACAGCGGAATCAGAAAGTTTGGATAGTAAAGAA 466

DB 1282 AspileLeuLysGlyLeuLysLysAspSerPheSerGlnGluSerSerProSerSerPro 1301

QY 467 AACCAACATACAGAAATAGAAATCCATGATGAGTCTGTACAAAAGATAAATTTTACCAA 526

DB 1302 SerAspLeuAlaLysLeuGluSerThrValLeuSerIleLeuGlu-----1316

QY 527 CATATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGTCTTGTAGTACTACTGAA 586

DB 1317 -----AlaGlnAlaSerThrLeuAlaAspGluLysSer-----Glu 1328

QY 587 AAAAGTACACAGTATTTGAACACGATCAGATCAGCAATGTGTAGTGGCAAAATGAA 646

DB 1329 LysLysThrGln-----ProHisGluValSerPro-----GluGlnProLysAsp 1343

QY 647 CGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACACAGTAACCTGTGACACGAG 706

DB 1344 GlnGluLysThrGlnSerLeu-----SerGluThrLeuGluIleThrIleSerGluGlu 1361

QY 707 CAGTTTAGTAATGTAACTGATTCG-----TCACCTCAAAATGAT 748

DB 1362 GluIleLysGluSerGlnGluGluArgLysAspThrPheLysLysAspSerGlnGlnAsp 1381

QY 749 GATCAGTGCACAGATAGTGAAGAAATAGAGACAATCAGACAGTTCCTCACAACCTGTA 808

DB 1382 IleProSerSerLysAspHisLysGluLysSerGluPheValAspIleThrArg 1401

QY 809 AAG-----CTTGCAAAATGCAGAGTACCGAAGATGAACACGCC-----850

DB 1402 ArgGluProTyrAspSerValGluGluSerSerGluSerGluAsnSerProValProGln 1421

QY 851 AGAAGAGCCAAAGCCACAGAGTGCAGCAAGTCTTGCCATCTCTGGGGAAGACTGTGCA 910

DB 1422 ArgLysArgArgThrSerValGlySerSerSerAspGluTyrLysGlnGluAspSer 1441

QY 911 AGTTGTCAGCAGATGAGATAGATAGCTGTGTGCCAAAGAGTCCATTTGATGATTTGGCTCT 970

DB 1442 GlnGlySerGlyGluGluGluAspPheIleArgLysGln---IleIleGluMetSerAla 1460

QY 971 GAGGATGTTGTACTGGGTCAAAAATGACAACTATTGATTAGACAAAGATGTCCTA 1030

DB 1461 AspGluAspAlaSerGlySerGluAspGluPheIleArgAsnGlnLeuLysGluIle 1480



QY 1031 GGAAATCTCTCCATTTGAGAGGAAGTGAACCCGAA----- 1069  
Db 1481 SerSerSerThrGluSerGlnLysLysGluGluThrLysGlyLysGlyLysLeuThrAla 1500  
QY 1070 -----TCACCGATGATGGATTAATCTTAA 1096  
Db 1501 GlyLysHisArgArgLeuThrArgLysSerSerThrSerileAspGluAspAlaGlyArg 1520  
QY 1097 AATAGTTGTCAAGACTCAGAACGAGATGAGAGACAAGTCACAGTTTGTATGATCAACAGAA 1156  
Db 1521 ArgHisSerTrpHisAspGluAspAspGlu-----AlaPheAspGluSerPro 1536  
QY 1157 GATGGTAGTTCTCCCAACAGCAATAAACCTTCA----- 1192  
Db 1537 GluLeuLysTyrArgGluThrLysSerSerGlnGluSerGluGluLeuValThrGlyGly 1556  
QY 1193 -----AGTTTCCCAAGCAAGACGCTGACATTGAATTTAGGAACCGTACTCT 1240  
Db 1557 GlyGlyLeuArgArgPheLysThrLleGluLeuAsnSerThrileAlaAspLysTyrSer 1576  
QY 1241 ACTAAGGCGGTGAAGTAGATTACATTTCCAAATT----- 1276  
Db 1577 AlaGluSerSerGlnLysLysThrSerLeuTyrPheAspGluGluProGluLeuGluMet 1596  
QY 1277 -----GAAGGAGGAGAGAGTGCACACT-----GGAATGAATGATTTAAAT 1315  
Db 1597 GluSerLeuThrAspSerProGluAspArgSerArgGlyGluGlySerSerLysHis 1616  
QY 1316 GCTAAA-----CTACCTGGAATATTTCTAGCCCTGAATGTAGATGTC 1357  
Db 1617 AlaSerSerPheThrProGlyThrSerProThrSerValSerSerLeuAspGluAspSer 1636  
QY 1358 AGAAATCTAAGCAACATGAGAAAGATCTTAA-----ATCACA 1399  
Db 1637 AspSerSerProSerHisLysLysGlyGluSerLysGlnArgLysAlaArgHisArg 1656  
QY 1400 GATCATTTGATGAGCTGCCAAGACAGACAGACA-----AGAAA 1441  
Db 1657 ProHisGlyProLeuLeuProThrileGluAspSerSerGluGluGluGluLeuArgGlu 1676  
QY 1442 GAACAGTGGGAACCAACATCAAGAACACAGAAAGAGATC----- 1483  
Db 1677 GluGluGluLeuLysGluGlnGlnLysGlnArgGluileGluGlnGlnArgLys 1696  
QY 1483 ----- 1483  
Db 1697 SerSerSerLysLysSerLysLysAspLysAspGluLeuArgAlaGlnArgArgGlu 1716  
QY 1484 ---CCTAAATACGTTCCACCTCACCTTCTCCAGATAAGAGTGGCTTGAACCTCCCAT 1540  
Db 1717 ArgProLysThrProProSerAsnLeuSerProileGluAsp-----AlaSerProThr 1734  
QY 1541 GAGGAGATGAGAAGATGCTCGGTGTGGATCGCGTCTCTCTGAGACCATCT--- 1597  
Db 1735 GluGluLeuArgGlnAlaAlaGluMetGlu-----GluLeuHisArgSerSerCys 1751  
QY 1598 GCCAATCACACAGTAATCTTCGGGTAGATCTTTTGGCAGCAGGAGAGTTCCCTAAACCT 1657  
Db 1752 SerGluTyrSerProSerileGluSerAsp-----ProGluGly 1764  
QY 1658 TTTCCACACATTAAGATTTTGGGAT---AACAGCATGTTAAATGCCCTTGTTC 1714  
Db 1765 PheGluLeuSerProGluLysLysileileGluValGlnLysValTyrLysLeuPro---Thr 1783  
QY 1715 GAACAAATTTGTCCAGTGGAGATGAGATGAGATGGTGAAGCACTGCGGGAGCGGTGG 1774  
Db 1784 AlaValSerLeuTyrSerProThrAspGluGln----- 1794  
QY 1775 GAGCTCATTCAGATGCACTTCTCAACAAATTTACACACCCCAAACTTCAAGATGCT 1834  
Db 1795 -----SerileMetGlnLysGluGlySerGlnLysAlaLeuLysSerAla 1809

QY 1835 ATTCTGAAATACAACTGTCATATTTCTAAGAAATGGGACTTTACAGCTTTG-----ATC 1888  
Db 1810 GluGluMetTyrGluGluMetMetHisLysThrHisLysTyrLysAlaPheProAlaAla 1829  
QY 1889 GATTCTGGGATAGGTACTTTGAAGAAGCAGAA-----GCTCAA 1927  
Db 1830 AsnGluArgAspGluValPheGluLysGluProLeuTyrGlyGlyMetLeuIleGluAsp 1849  
QY 1928 CATTTATATCAGTCCATCTTGCCTGAT 1954  
Db 1850 TyrileTyrGluSerLeuValGluAsp 1858  
RESULT 5  
PCLO MOUSE STANDARD; PRT; 5038 AA.  
ID PCLO\_MOUSE Q9QYX7; Q9QYX6; Q9QZJ0;  
AC Q9QYX7; Q9QYX6; Q9QZJ0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain-derived HLMN protein).  
GN PCLO OR ACZ.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, ALTERNATIVE SPLICING, TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.  
RC TISSUE=Brain;  
RX MEDLINE=99439764; PubMed=10508862;  
RA Wang X., Kibechull M., Laue M.M., Lichte B., Petrasch-Parwez E., Kilimann M.W.;  
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active zones, shares homology regions with rim and bassoon and binds profilin.";  
RL J. Cell Biol. 147:151-162 (1999).  
RN [2]  
RP REVISIONS.  
RA Kilimann M.W.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 4502-4682 FROM N.A.  
RC TISSUE=Brain;  
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP INTERACTION WITH RIMS2.  
RX MEDLINE=22384373; PubMed=12401793;  
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M., Sasaki T., Tajima N., Iwanaga T., Seino S.;  
RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of CAMP-GEFII.Rim2.Piccolo complex in CAMP-dependent exocytosis.";  
RL J. Biol. Chem. 277:50497-50502 (2002).  
CC -!- FUNCTION: May act as a scaffolding protein involved in the organization of synaptic active zones and in synaptic vesicle trafficking.  
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.  
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of synaptic junctions.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9QYX7-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9QYX7-2; Sequence=VSP\_003928, VSP\_003929;  
CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in stomach. Not detected in other tissues analyzed including adrenal gland, testis and pancreas.  
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and phospholipids. Calcium binds with low affinity but with high specificity and induces a large conformational change.  
CC -!- SIMILARITY: Contains 2 C2 domains.







QY 2477 -----TTCACCTGAGGCTG-----GATCACATGATGCTCAATT 2512  
 Db 1817 ArgLyAspTyrThrAlaLeuThrLysPheLeuProProlYsHisGluTyrValLeuAla 1836  
 QY 2513 ATCCACAGGTACTGAG--CAGTACATGATACACAGGCTATGCTGAGACATATCGTGG 2569  
 Db 1837 ValArgMetThrSerIleGlnCysLeuTyrGlnTyrTyrLeuAspHisLeuThrGlu 1856  
 QY 2570 TCCCGAGGACCCAGAGTGGGAGTGAAGGACGACTGGCGGCGCTGCACCTG 2624  
 Db 1856 yValGlyAsnAsnSerGluGlyGlyArgGly-----LysAlaGlyAlaLysLeu 1872  
 RESULT 7  
 ATRX HUMAN  
 ID ATRX HUMAN STANDARD; PRT; 2492 AA.  
 AC P46100; P51068; Q15886; Q9H021; Q9NTS3;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked  
 DE nuclear protein) (XNP) (Znf-HX).  
 GN ATRX OR RAD54L OR XH2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANTS SER-596;  
 RP GLU-740 AND SER-1860, AND VARIANTS ATR-X.  
 RX MEDLINE-97123494; PubMed-8968741;  
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,  
 RA Gibbons R.J.;  
 RA "ATRX encodes a novel member of the SNF2 family of proteins: mutations  
 RT point to a common mechanism underlying the ATR-X syndrome.";  
 RL Hum. Mol. Genet. 5:1899-1907(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).  
 RX MEDLINE-97386582; PubMed-9244431;  
 RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,  
 RA Colleaux L., Schwartz C., Fontes M.;  
 RA "Determination of the genomic structure of the XNP/ATRX gene encoding  
 RT a potential zinc finger helicase.";  
 RL Genomics 43:149-155(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANTS SER-596 AND GLU-740.  
 RX MEDLINE-22763540; PubMed-12777533;  
 RA Kitano T., Schwarz C., Nickel B., Paabo S.;  
 RA "Gene diversity patterns at 10 X-chromosomal loci in humans and  
 RT chimpanzees.";  
 RL Mol. Biol. Evol. 20:1281-1289(2003).  
 RN [4]  
 RP SEQUENCE OF 860-2492 FROM N.A.  
 RX MEDLINE-95179111; PubMed-7874112;  
 RA Stayton C.L., Dabovic B., Gullisano M., Gez J., Broccoli V.,  
 RA Giovanazzi S., Boscosasco M., Monaco L., Rastan S., Boncinelli E.,  
 RA Bianchi M.E., Consalez G.G.;  
 RA "Cloning and characterization of a new human Xq13 gene, encoding a  
 RT putative helicase.";  
 RL Hum. Mol. Genet. 3:1957-1964(1994).  
 RN [5]  
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE-94214473; PubMed-8162050;  
 RA Gez J., Pollard H., Consalez G., Villard L., Stayton C.L.,  
 RA Millaesau P., Khrestchatsky M., Fontes M.;  
 RA "Cloning and expression of the murine homologue of a putative human  
 RT X-linked nuclear protein gene closely linked to PGK1 in Xq13.3.";  
 RL Hum. Mol. Genet. 3:39-44(1994).  
 RN [6]  
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.  
 RX MEDLINE-95211835; PubMed-7697714;  
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
 RA "Mutations in a putative global transcriptional regulator cause X-  
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";  
 RL Cell 80:837-845(1995).  
 RN [7]  
 RP SEQUENCE OF 1375-2492 FROM N.A.  
 RA Pearce A., Chapman J.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP EZH2 BINDING.  
 RX MEDLINE-98167853; PubMed-9499421;  
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,  
 RA Colleaux L.;  
 RA "Specific interaction between the XNP/ATRX gene product and the SET  
 RT domain of the human EZH2 protein.";  
 RL Hum. Mol. Genet. 7:679-684(1998).  
 RN [9]  
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE-20040663; PubMed-10570185;  
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Bickmore W.A., Pombu A., Turley H., Gatter K., Picketts D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RA "Localization of a putative transcriptional regulator (ATRX) at  
 RT pericentromeric heterochromatin and the short arms of acrocentric  
 RT chromosomes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 RN [10]  
 RP DISEASE.  
 RX MEDLINE-20213147; PubMed-10751095;  
 RA Villard L., Fontes M., Ades L.C., Gez J.;  
 RA "Identification of a mutation in the XNP/ATRX gene in a family  
 RT reported as Smith-Fineman-Myers syndrome.";  
 RL Am. J. Med. Genet. 91:83-85(2000).  
 RN [11]  
 RP DISEASE.  
 RX MEDLINE-22804448; PubMed-12858175;  
 RA Gibbons R.J., Pellagatti A., Garrick D., Wood W.G., Malik N.,  
 RA Ayub H., Langford C., Boulwood J., Wainscoat J.S., Higgs D.R.;  
 RA "Identification of acquired somatic mutations in the gene encoding  
 RT chromatin-remodeling factor ATRX in the alpha-thalassemia  
 RT myelodysplasia syndrome (ATMDS).";  
 RL Nat. Genet. 34:446-449(2003).  
 RN [12]  
 RP VARIANT ATR-X SER-1713.  
 RX MEDLINE-97196774; PubMed-9043863;  
 RA Villard L., Lacombe D., Fontes M.;  
 RA "A point mutation in the XNP gene, associated with an ATR-X phenotype  
 RT without alpha-thalassemia.";  
 RL Eur. J. Hum. Genet. 4:316-320(1996).  
 RN [13]  
 RP VARIANT JM GLN-2131.  
 RX MEDLINE-96224392; PubMed-8630485;  
 RA Villard L., Gez J., Mattei J.-F., Fontes M., Saugier-Verber P.,  
 RA Munnich A., Lyonnet S.;  
 RA "XNP mutation in a large family with Juberg-Marsidi syndrome.";  
 RL Nat. Genet. 12:359-360(1996).  
 RN [14]  
 RP VARIANTS ATR-X.  
 RX MEDLINE-97467722; PubMed-9326931;  
 RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,  
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,  
 RA Higgs D.R.;  
 RA "Mutations in transcriptional regulator ATRX establish the functional  
 RT significance of a PHD-like domain.";  
 RL Nat. Genet. 17:146-148(1997).  
 RN [15]  
 RP VARIANT ATR-X LEU-246.  
 RX MEDLINE-20123062; PubMed-10660327;  
 RA Fichera M., Romano C., Castiglia L., Failla P., Ruberto C., Amata S.,  
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;  
 RA "New mutations in XNP/ATRX gene: a further contribution to  
 RT genotype/phenotype relationship in ATR/X syndrome.";  
 RL Hum. Mutat. 12:214-214(1998).





CC EMBL; AB102643; BAC81112.1; --  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR000330; SNF2\_N.  
DR Pfam; PF00271; Helicase\_C; 1.  
DR Pfam; PF00176; SNF2\_N; 1.  
DR SMART; SM00487; DEXDC; 1.  
DR SMART; SM00490; HELICC; 1.  
KW DNA repair; Nuclear protein; DNA-binding; Helicase; ATP-binding;  
KW Zinc-finger.  
FT ZN\_FING 220 268 PHD-TYPE.  
FT NP\_BIND 1594 1601 ATP (POTENTIAL).  
FT SITE 1719 1722 DEGH BOX.  
FT DOMAIN 745 750 POLY-SER.  
FT DOMAIN 1151 1156 POLY-SER.  
FT DOMAIN 1166 1169 POLY-LYS.  
FT DOMAIN 1202 1206 POLY-SER.  
FT DOMAIN 1259 1266 POLY-ASP.  
FT DOMAIN 1443 1466 POLY-GLU.  
FT DOMAIN 1499 1502 POLY-GLU.  
FT DOMAIN 1929 1939 POLY-LYS.  
FT DOMAIN 1941 1948 POLY-SER.  
FT DOMAIN 2222 2225 POLY-LYS.  
FT DOMAIN 2262 2265 POLY-GLU.  
FT DOMAIN 2420 2425 POLY-GLN.  
SQ SEQUENCE 2492 AA; 282614 MW; 48EC97106D0CEF20 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.00582 Length: 2492  
Score: 161.00 Matches: 175  
Percent Similarity: 36.09% Conservative: 135  
Best Local Similarity: 20.37% Mismatches: 314  
Query Match: 2.23% Indels: 235  
DB: 1 Gaps: 47  
  
US-09-302-812-3 (1-4069) x ATRX\_PONPY (1-2492)  
QY 389 AAACAAAGACATTATACCAAGTGGTGGACATAAGGAATCAAGACAGCGGAATCAGAA 448  
DB 1136 GluArgAsnLeuSerSerLysArgAsnThrLysGluLeGlnSerGlySerSer 1155  
QY 449 AGTTGGATAGTAAGAAAC-----AACATACAGAAATAGATCCCATGATGTTCT 502  
DB 1156 Ser---AspAlaGluGluSerSerGluAspAsnLysLysLysGlnArgThrSerSer 1174  
QY 503 GTACAAAGATACCTTTTACCACATATGTAGAAATATTAGTAAATGTTTCTCAGCTA 562  
DB 1175 LysLysLysAlaValleVallyLysLysLysArgAsnSerLeuArgThrSerThrLys 1194  
QY 563 AGTCTTGATAAGTCACTCACTGAAAAAGTACACAGTATTTTGAACCGCATCAGACTGCA 622  
DB 1195 ArgLysGlnAlaAspIleThrSerSerSerSerSerSerSerSerSerSerSerSer 1214  
QY 623 GCAATGTGTAAGTGGCAAAATGAGGAAACACACGAGCAGCTTTTGAAGTGAACCT 682  
DB 1215 SerIle-----GlyGluGlySerSerSerSerSerSerSerSerSerSerSerSer 1228  
QY 683 CAAACAGTAACCTGGTACCAGAGCAG-----TTTAGTAATGCTAACATTGATCGG 733  
DB 1229 ValThrGluAsnLeuValleuSerSerHisThrGlyPheCysGlnSerSerGlyAspGlu 1248  
QY 734 TCA-----CCTCAAAATGATGATCAGTGCACACAGATAGTGA 772  
DB 1249 AlaLeuSerLysSerValProValThrValAspAspAspAspAspAspAspProGlu 1268  
QY 773 GAGATAGACACAAATCAAGTTTCTCACAAGTGTAAAG-----CTTGCAATGCAAG 826  
DB 1269 AsnArgIleAlaLysLysMetLeuLeuGluGluGluLeuLysAlaAsnLeuSerSer 1288  
QY 827 CAGACTACGGAAGTACACGCGCAGAGAGCCAAAGCCACCAGAGTGCAGCAAGTCT 886  
DB 1289 AspGlySerSerAspAspGluProGluGluGlyLysLysArgThrGlyLysGlnAsnGlu 1308

QY 887 TGCATCTCTGGGGAAGACTGTGCAAGTGTTCAGCAAGATGAGATAGACGTGGTGCCAAAG 946  
DB 1309 GluAsnProGlyAspGluGluAlaLysAsnGlnValAsnSer----- 1322  
QY 947 AGTCCATTGTCAGATGTTGGCTCTGAGAGTGTGGTACTGGTCAAAAATAGC----- 1000  
DB 1323 -----GluSerAspSerAspSerGluGlu-----SerLysLysProArgTyr 1336  
QY 1001 ---AACAAATTGATTAGCAAGAAAGTTGCTAGAGAAATTCCTCTCCA---TTTGAGAG 1054  
DB 1337 ArgHisArgLeuLeuArgHisLysLysLeuThrValSerAspGlyGluSerGlyGluLys 1356  
QY 1055 GAAAGTGAACCCGAATCACCATGCGATGTG-----GATAATCTTAAATAAGTTGT 1105  
DB 1357 LysThrLysProLysGluHisLysGluValLysGlyArgAsnArgArgLysValSerSer 1376  
QY 1106 CAAGACTCAGAAAGCAGATGAGGAGACCAAGTCCAGTTTGTGATGATCAACAGAGATG 1165  
DB 1377 GluAspSerGluAspSerAspPheGlnGluSerGlyValSerGluGluVal-----Ser 1394  
QY 1166 TCCTCCCAACACAGCAAAATAAACCTTCAAGGTTCCAAAGACAGACGCTGACATTGAAT 1225  
DB 1395 GluSerGluAspGluArgProArgThrArgSerAlaLysLysAlaGluLeuGlu 1414  
QY 1226 AGGAAACGG---TACTCTACTAAGCGGTGCAAGTTAGATTACATTTCCTCAATTTGAAG 1282  
DB 1415 AsnGlnArgSerTyrLysGlnLysLysArgArgGlyLysValGlnGluAspSer 1434  
QY 1283 GGA----- 1285  
DB 1435 SerSerGluAsnLysSerAsnSerGluGluGluGluGluLysGluGluGluGlu 1454  
QY 1286 -----GAGAGTCGCACCTGGAATGAATGATTTAAATGCTAAACTACTCT 1327  
DB 1455 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1473  
QY 1328 GGAATATTCTAGCCTCAATGTAGATGCAAGAAATCTTAGCAACATGGAATA----- 1381  
DB 1474 Gly-----LysGlyArgLysLysLysLysLysLysLysLysLysLysLys 1484  
QY 1382 AAGGATTTCTAAATC---ACAGATCATTTGATGACACTGCCCAAGACAGACAGACAGA 1438  
DB 1485 LysAspAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1504  
QY 1439 AAAGAACAGTGGGAACCAACATCAAGAACAGAAAGAGATCCCTAAATACCTTCCA 1498  
DB 1505 LysArgLysAlaGluArgGluArg-----GluLysLeuArgGluValIleGlu 1522  
QY 1499 ---CCTCACCTTCTCCAGATAAGAGTGGTGGAACTGCCAT----- 1540  
DB 1523 IleGluAspAlaSerProThrLysCysProIleThrLysLysLysLysLysLysLys 1542  
QY 1541 GAGGAGATGAGAGAAATGCTGGGTGGGATCCCGCTCTCTTGGAGACCATCTGCC 1600  
DB 1543 GluGluThrLysGlu-----ProLeuValGlnValHisArg 1554  
QY 1601 AATCAGACAGTAATCTATTCGGGTAGATCTTTTGGCAGGAGAGATTCCTTAACTTTT 1660  
DB 1555 AsnMetValIleLysLeu-----LysPro----- 1562  
QY 1661 CCAACACATTAATAAGAT-----TTGTGGGATAACAGCATGTTAAAAATGCCT 1708  
DB 1563 -----HisGlnValAspGlyValGlnPheMetTrpAsp----- 1573  
QY 1709 TGTTCAGAACAAAATTTGTACCCAGTGAAGATGAGATGGTGAGCAACTCGCGGAGC 1768  
DB 1574 CysCysCysGluSerVal-----LysLysThrLysLysSerProGlySer 1588  
QY 1769 CGGTGGGAGCTATTTCAGACTCCTCTTCAACAATTTACACGCCCCCAAACTTGAG 1828  
DB 1589 GlyCysIleLeuAlaHisCysMetGlyLeuGlyLys-----ThrLeuGln 1603





FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Alignment Scores:  
Pred. No.: 0.00637 Length: 1957  
Score: 160.00 Matches: 182  
Percent Similarity: 35.63% Conservative: 149  
Best Local Similarity: 19.59% Mismatches: 343  
Query Match: 2.22% Indels: 255  
DB: 1 Gaps: 39

US-09-302-812-3 (1-4069) x SPOF\_SCHPO (1-1957)

```
QY 242 TCGAGCGCGGAGCTTTCCAGCAGCAGCGCGCTCTCGACCCCAAGACGCTCAC 301
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db SerThrGlnSerSerLeuGlnSerLeuGlnArgAspIleLeuAsnGluLysLysHis 527

QY 302 GTGCAGTTCAGGGTCCCGCTCTCGCAGCGCTCGGCCCGCAGCGGAGCAGCAC 361
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GluValTyrGlu-----SerGlnLeuAsnGluLeu 537

QY 362 AGAGCAGCGCCACCTCGCTTGTTCACAAACAAAGACTATTACCAGTTGGATGACACT 421
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LysGlyGluLeuGlnThrGluLysSerAsnSerGluHisLeuSerGlnLeuSerThr 557

QY 422 AAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAGAAACAAACAATCAAGA 481
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----LeuAlaAlaGluLysGluAla-----AlaValAlaThrAsnAsnGluLeu 572

QY 482 ATAGATCATGATGAGTTCTGTACAAAGATAACTTTTACCACATAATGTAGAAAAA 541
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db SerGluSerLysAsnSerLeuGlnThrLeuGlnThrLeuGlnThrLeuGlnThrLeu 592

QY 542 TTAGTAATATGTTTCTCAGCTA-----AGTCTTCAT----- 571
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----SerValMetGlnLeuLysGluAsnGluGlnAsnPheSerSerLeuAspThrSer 610

QY 572 ---AAGTCACTCACTCAAAAAGTACACAGTATTTGAACAGCATCAGACTGCACCAATG 628
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db PheLysLysLeuAsnGluSerHisGlnGluLeuGluAsnAsnHisGlnThrIle----- 628

QY 629 TGTAAAGTGCAAAATGAAGGGAAACACACGGAGAGCTTTTGAAGAGTGAACCTCAACA 688
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----ThrLysGlnLeuLysAspThrSerSerLysLeu 639

QY 689 GTAACCTGGTACCAGACAGTTAGTAATGCTAACTGATCGCTCACTCAAAATGAT 748
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GlnGlnLeu-----GlnLeuGluArgAlaAsnPheGluGlnLys----- 652

QY 749 GATCAGTGACACAGATAGTGAAGAGATAGAGACAATCAACAGTTTCTCAACTGTGA 808
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----GluSerThrLeuSerAspGluAsnAsnAspLeuArgThrLysLeuLysLeu 670

QY 809 AAGCTTGCAAAATGCAAG----- 826
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GluGluSerAsnLysSerLeuLysLysLysGlnGluAspValAspSerLeuGluLysAsn 690

QY 827 ---CAGACTACGGAAGATGAACACGCGACA----- 853
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db IleGlnThrLeuLysGluAspLeuArgLysSerGluGluAlaLeuArgPheSerLysLeu 710

QY 854 GAAGCAAAAGCCACCAAGATGTCAGCAAGTCTTGGCCATCTCGGGGAGACACTGTGCAAGT 913
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GluAlaLysAsnLeuArgGluValIleAspAsnLeuLysGlyLysHisGlnThrLeuGlu 730

QY 914 TGTCAAGATGATGATAGACGTGGTGCCAAAGATTCATTGTCAGATCTTGCTCTGAG 973
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----HisSerSerLeuSerAsp----- 742

QY 974 GATGTTGTGCTGGTCAAAAATGACAAATATTGATTAGACAAGATTTGCCCTAGGA 1033
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----AlaLysAsnThrAsnAlaIleLeuSerSerGlu----- 753
```

```
QY 1034 AATTCTCCTCCATTGTGAGAAGAAAGTGAACCCGAATCACCGATGATGTGATTAATTCT 1093
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----LeuThrLysSerSerGluAspValLysArgLeuThrAla-----AsnVal 768

QY 1094 AAAATAGTTGTCAAGACTCAGAAGCAGATGAGAGAGCAAGTCCAGGTTTGTATGAACAA 1153
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GluThrLeuThrGlnAspSerLysAla-MetLysGlnSerPheThrSerLeuValAsnSe 788

QY 1154 CAAGATGTGATGTTCTCT-----CCCAACAGCAAAATAAACCTTCAAGG 1195
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db rTyrGlnSerIleSerAsnLeuTyrHisGluLeuArgAspAspHisValAsnMetGln-- 807

QY 1196 TTCCAAGCAAGAGCGCTG-----ACATT 1219
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -SerGlnAsnAsnThrLeuLeuGluSerGluSerLysLeuLysThrAspCysGluAsnLe 827

QY 1220 GAATTTAGGAACGCTACTTACTTAAGCGCGTGAAGTTAGATTACATTTCCTCAATTTGAA 1279
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db uThrGlnGlnAsnMetThrLeuIleAspAsnValGln--LysLeuMetHisLysHisVal 846

QY 1280 CGAGGAGAGAGTCGCACCTGGA---ATGAATGATTTAAATGCTAAACTACCTGGAAATATT 1336
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db AsnGlnGluSerLysValSerGluLeuLysGluValAsnGlyLysLeuSerLeuAspLeu 866

QY 1337 -----TCTAGCCTGAATGTAGAATGCGAAATTTCTAAGCAACATCGGAAAAAG 1384
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LysAsnLeuArgSerSerLeuAsnValAlaIleSerAspAsnAspGln----- 882

QY 1385 GATTCTAAATACACAGATCATTTTGATGAGACTGCCAAAGCAGAGGAGCAAGAAAAAGAA 1444
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----IleLeuThrGlnLeuAlaGluLeuSerLysAsnTyrAspSerLeuGluGln 899

QY 1445 CAGTGGGAAACCAACATCAAGAACAGAAAGAGATCCCTAAATACGTTCACCTCAC 1504
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----GluSerAlaGlnLeuAsnSerGlyLeuLys-----Ser 910

QY 1505 CTTTCTCCAGATAGAGTGGCTTGGAACTCCATTGAGGAGATCAGAGAATGCTCGG 1564
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LeuGluAlaGluLysGlnLeuLeuHisThrGluAsnGluGluLeuHis----- 926

QY 1565 TGTGGGATCGCGCTCTCTCTGAGACCATCTGCCAATCACACAGTAATCTATTCGGGTA 1624
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----IleArgLeuAspLysLeu-----ThrGlyLysLeuLysIle 938

QY 1625 GATCTTTTCGAGCAGGAGAGAGTTCCTTAAACCTTTTCCACACATTAAGATTGTGG 1684
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db GluGluSerLysSerSerAspLeuGlyLysLysLeuThrAlaArgGlnGluLysSer 958

QY 1685 GATAACAGCATGTTAAATGCTTGTTCAGAACAAAATTTGTACCCAGTGAAGATGAG 1744
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db AsnLeuLysGluGluAsnMetSerGlnSer--GlnAlaIleThrSerValLysSerLys 977

QY 1745 AATGTGTAGCGCA---ACTGCGGGGCGCGGTGGGAGCTCATTCAGACTGCTTCAC 1801
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LeuAspGluThrLeuSerLysSerLysLeuGluAlaAspIleGluHisLeuLysAsn 997

QY 1802 AAATTTACAGACCCCAAAACTTGAGGATCTATTCTGAATACAAATGTGCATATTCT 1861
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db LysValSerGluValGluValArgAsnAlaLeuLeuAlaSerAsn----- 1013

QY 1862 AAGAAATCGGACTTTACAGCTTTGATCGAT----- 1891
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db -----GluArgLeuMetAspAspLeuLysAsnAsnGlyLysLeuAla 1028

QY 1892 TTCTGGGTAAGGTACTTGAAGACAGAGCTCAACATTATATCATCTCCATCTTGCT 1951
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db SerLeuGlnThrGluIleGluLysLysArgAlaGluAsn----- 1041

QY 1952 GATATGGTGAATAATGCACTCTGCTGCCAATATTTGCAACCCCAATACCACTCTG 2011
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db AsnAspLeuGlnSerLysLeuSerValValSerGluTyrGluAsnLeuLeuLys 1061

QY 2012 AAACAGAGATGAATCATTTCCATCACAAATGTGCGAGGAACAGATTGCCAGTCTTTAGCT 2071
```

```

Db 1062 SerSerGlnThrAsnLysSerLeuGluAspLysThrAsnGlnLeuLysTyrIleGluLys 1081
QY 2072 AATGCT-----TTCCTCTCCATTTCCACGAAATCTAGATGAATCGAGTAT 2125
Db 1082 AsnValGlnLysLeuLeuAspGluLysAspGlnArgAsnValGluLeuGluLeuThr 1101
QY 2126 TCTAGTTACCCACATTAATCTCAATCGATGTTTTCAGGAGCTTCATCAAGAAACCG 2185
Db 1102 SerLysTyr-----GlyLysLeuGlyGluGluAsn 1111
QY 2186 GAGAACTTAAACGCTCTTCTGCTACTTTAGAGAGTCACAGAGAAATGCTACTGGG 2245
Db 1112 AlaGlnIleLysAspGluLeuLeuAlaLeuArgLysSerLysLysGlnHisAspLeu 1131
QY 2246 TTGGTGATTTCAAGACAGAGTCCTTGAAGATTTTCCAGAAATGGGAAAGATGTGAAAA 2305
Db 1132 CysAlaAsnPheValAspAspLeuLysGluLysSerAspAlaLeuGlnLeuThrAsn 1151
QY 2306 CCCTTGACAGATTGCTATGCTTACGAGGTACCATAGAGAAATGGCCAGGCATG 2365
Db 1152 GluLysAsnGluLeuIleValSerLeuGluGlnSer---AsnSerAsnAsnGluAlaLeu 1170
QY 2366 CTACAG-----GTGGATTTTGAATATCGTTTGTGGAGGTGGTAAACAGTCGACGA 2419
Db 1171 ValGluGluArgSerAspLeuAlaAsnArg----- 1180
QY 2420 CTTGTGCAAGAGAAATCCGCTTTTAAATCAATCTGAGTTGATTAATTCACGGCTTTC 2479
Db 1181 ---LeuSerAspMetLysLysSerLeuSerAspSerAspAsnValIleSerValIleArg 1199
QY 2480 ACTGAGGTCTGTCATCATCATGATGCTAATATATCATCAGTACTGACGAGTACAGTAA 2539
Db 1200 SerAspLeuValArgValAsnAspGluLeuAspThrLeuLysLysAspLysAspSerLeu 1219
QY 2540 TACACAGGCTATGTCAGACATATCGTTGTGTCGCGGAGCCACGAGATGGAGTGAAGG 2599
Db 1220 SerThrGlnTyrSerGluValCysGln-----AspArg 1230
QY 2600 GACGAC-----TGCAGCGCGCGC----- 2617
Db 1231 AspAspLeuAspSerLeuLysGlyCysGluGluSerPheAsnLysTyrAlaValSer 1250
QY 2618 -----TGCACCT----- 2623
Db 1251 LeuArgGluLeuCysThrLysSerGluIleAspValProValSerGluIleLeuAspAsp 1270
QY 2624 -----GAGATCGTTGCCATCGATGCTCTTCACTTC 2653
Db 1271 AsnPheValPheAsnAlaGlyAsnPheSerGluLeuSerArgLeuThrValLeuSerLeu 1290
QY 2654 AGACGTACTCTCATCGTTT 2674
Db 1291 GluAsnTyrIleuAspAlaPhe 1297

```

## RESULT 10

```

BMS1_YEAST
ID_BMS1_YEAST STANDARD; PRT; 1183 AA.
AC Q08965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosome biogenesis protein BMS1.
GN BMS1 OR YPL217C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,

```

```

RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipalo T., Dubois E., Dueterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Labkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Maerche R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Urestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
RN [2]
RP FUNCTION
RX MEDLINE=21449426; PubMed=11565748;
RT Wegierski T., Billy E., Nasr F., Filipowicz W.;
RT "Bms1p, a G-domain-containing protein, associates with Rclp and is
RT required for 18S rRNA biogenesis in yeast.";
RL RNA 7:1254-1267(2001).
CC -i- FUNCTION: May act as a molecular switch during maturation of the
CC 40S ribosomal subunit in the nucleolus. The depletion of BMS1
CC interferes with processing of the 35S pre-rRNA at sites A0, A1,
CC and A2, and the formation of 40S subunits.
CC -i- SUBUNIT: Associates with RCL1.
CC -i- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -i- SIMILARITY: Belongs to the BMS1 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73573; CAA37932.1; -
CC PIR; S65236; S65236.
CC SGD; S0006138; BMS1.
CC GO; GO:0005737; Cytoplasm; IDA.
CC GO; GO:0005730; Nucleolus; IDA.
CC GO; GO:0006365; P:35S primary transcript processing; IDA.
CC GO; GO:0042255; P:ribosome assembly; IDA.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR007034; DUF663.
CC Pfam; PF04950; DUF663; 1.
CC SMART; SM00382; AAA; 1.
KW Ribosome biogenesis; Nuclear protein; ATP-binding.
FT NP_BIND 76 83 ATP (POTENTIAL).
SQ SEQUENCE 1183 AA; 135570 MW; 9A337F1E0B0F21D CRC64;

```

## Alignment Scores:

```

Pred. No.: 0.00603 Length: 1183
Score: 159.50 Matches: 156
Percent Similarity: 31.89% Conservative: 116
Best Local Similarity: 18.29% Mismatches: 274
Query Match: 2.21% Indels: 307
DB: 1 Gaps: 36

```

US-09-302-812-3 (1-4069) x BMS1\_YEAST (1-1183)

```

QY 263 AGCAGCAGAGCGCGTCTCTCGACCCCAAGGAC----- 295
Db 341 ThrArgArgArgLysArgLeuAspLysAspLeuLeuTyrAlaProMetSerAsp 360
QY 296 -----GCTCAGTCGAGTTCAGGTCACCGTCC 325
Db 361 ValGlyGlyValLeuMetAspLysAspAlaValTyrIleAspIleGlyLysLysAsnGlu 380

```

326 TCGCCAGCCTCGTCCAGGCGAG---GCGGACACAGCAGAGGCGCGCCACCTCGCTT 382  
Db |||||  
381 GluProSerPheValProGlyGlnGluArgGlyGluGlyGluLysLeuMetThrGlyLeu 400  
Qy |||||  
383 GTTTTCAACAAAGACTATTACAGCTGGTGGATGACACTAAGGAATC----- 430  
Db |||||  
401 GlnSerValGluGlnSerIleAlaGluLysPheAspGlyValGlyLeuGlnLeuPheSer 420  
Qy |||||  
431 -----AACACAGCGGAATCAGAAAGTTTGGATAGTAAAGAAACAAAC 472  
Db |||||  
421 AsnGlyThrGluLeuHisGluValAlaAspHisGluGlyMetAsp---ValGluSerGly 439  
Qy |||||  
473 AATACAGAAATAGAA-----TCCATGATGAGTCTGTGTACAAAGAT 514  
Db |||||  
440 GluGluSerIleGluAspGluGlyLysSerLysGlyArgThrSerLeuArgLysPro 459  
Qy |||||  
515 AACTTTTCAACAAATAGTA---GAAAAATAGTAATGTTTCTCAGCTAAGTCTTGAT 571  
Db |||||  
460 ArgIleTyrGlyLysProValGlnGluGlnAspAlaAspIleAspAsnLeuProSerAsp 479  
Qy |||||  
572 AAGTCACTCACTGAAAAAGTACACAGTATTGAAACACAGCATCAGACTGCAGCAATGTGT 631  
Db |||||  
480 GluGlu-----ProTyrThrAsnAsp----- 486  
Qy |||||  
632 AAGTGGCAAAATGAAGGAAACACACAGGAGCGCTTTTGGAAAGTGAACCTCAACAGTA 691  
Db |||||  
487 -----AspAspValGlnAspSerGluProArgMetVal 497  
Qy |||||  
692 ACCCTGGTACACAGCAGCTTAGTAACTGCTAACATTGCTGCTCACCCTCAAAATCATGAT 751  
Db |||||  
498 GluIle-----AspPheAsnAsnThrGlyGluGlnGlyAlaGluLysLeuAlaLeu 514  
Qy |||||  
752 CACACTGACACAGATAGTGAAGAGAAATAGAGCAATCAACAGTTTCTCACAACCTGTAAG 811  
Db |||||  
515 GluThrAspSerGluPheGluSerGluAspGluPheSerTyrGluArgThrAla--- 533  
Qy |||||  
812 CTTGCAAAATGCAACGACAGCTACGGAAGATGAACACGCGC----- 850  
Db |||||  
534 ---AlaAsnLysLeuLysLysThrGluSerLysLysArgThrTyrAsnIleGlyLysLeu 552  
Qy |||||  
851 -----AGAGAGCCAAAGCCACACAGAGTGCACCAAGTCTTGCCATCTGGGGAAGAC 904  
Db |||||  
553 IleTyrMetAspAsnIleSerProGluGluCysIleArgArgTyrArg---GlyGluAsp 571  
Qy |||||  
905 TGTGCAAGTTGTCAACAGATGAGATAGACGTGGTGCACAGAGTCCATTGTGCAGATGTT 964  
Db |||||  
572 AspAspSer-----LysAspGluSerAspIle----- 580  
Qy |||||  
965 GGCTCTGAGGATGTT-----GGTACTGGGTCAAAAAAT 997  
Db |||||  
581 ---GluGluAspValAspAspAspPhePheArgLysLysAspGlyThrValThrLysGlu 599  
Qy |||||  
998 GACACAA-----TTGATTAGACAGAAAGTTCGTAGGAAATCTCTCCATTTGAG 1051  
Db |||||  
600 GlyAsnLysAspHisAlaValAspLeuGluLysPheValProTyrPheAspThrPheGlu 619  
Qy |||||  
1052 AAGGAAAGTGAACCGGAATCACCGATGATGTG----- 1084  
Db |||||  
620 LysLeuAlaLysLysThrLysSerValAspAlaIleLysGluArgPheLeuGlyAlaGly 639  
Qy |||||  
1085 -----GATAATTCTAAAATAGTCTCAAGACTCAGAGCAGATGAGAGCAAGT 1135  
Db |||||  
640 IleLeuGlyAsnAspAsnLysThrLysSerAspSerAsnGluGlyGlyGluLeuTyr 659  
Qy |||||  
1136 CCAGGTTTTCATCAACAAAGATGTTAGTCTCTCCCAACAGCAAAATAAACCCTTCAAGG 1195  
Db |||||  
660 GlyAspPheGluAspLeuGluAspGlyAsnProSerGluGlnAlaGluAspAsnSerAsp 679  
Qy |||||  
1196 TTCCAGCAGCAGACGCTGAC----- 1216  
Db |||||  
680 LysGluSerGluAspGluAspGluAsnGluAspThrAsnGlyAspAspAsnSerPhe 699  
Qy |||||  
1217 -----ATTGAATTTAGGAAACGGTAC 1237

700 ThrAsnPheAspAlaGluGluLysLysAspLeuThrMetGluGlnGluAspGluMetAsn 719  
Qy |||||  
1238 TCTACTAAGCGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGGAGAGTCCGCACT 1297  
Db |||||  
720 AlaAlaLysLysGluLysLeuArgAlaGlnPheGluIleGluGluGlyGluAsnPheLys 739  
Qy |||||  
1298 GGAATGAATGATTTAAAT-----GCTAAACTACCT 1327  
Db |||||  
740 GluAspAspGluAsnAsnGluTyrAspThrTyrGluLeuGlnLysAlaLysIleSer 759  
Qy |||||  
1328 GGAATATTCTAGCTGAAATGAGATGC-----AGAAATCTT 1366  
Db |||||  
760 LysGlnLeuGluIleAsnAsnIleGluTyrGlnGluMetThrProGluGlnArgGlnArg 779  
Qy |||||  
1367 AAGCAACATCGAAAAAGGATTTCAAATCACAGATCATTTGTATGAGACTGCC- 1420  
Db |||||  
780 IleGluGlyPheLysAlaGlySerTyrValArgIleValPheGluLysValProMetGlu 799  
Qy |||||  
1420 ----- 1420  
Db |||||  
800 PheValLysAsnPheAsnProLysPheProIleValMetGlyGlyLeuLeuProThrGlu 819  
Qy |||||  
1421 -----AAAGCAGAGCAGACAGAAAGAAAGAAACAGTGGGAA- 1453  
Db |||||  
820 IleLysPheGlyIleValLysAlaArgLeuArgArg---HisArgTyrHisLysLysIle 838  
Qy |||||  
1453 ----- 1453  
Db |||||  
839 LeuLysThrAsnAspProLeuValLeuSerLeuGlyTyrArgArgPheGlnThrLeuPro 858  
Qy |||||  
1454 -----ACCAACATCAAAAGCAGAAAGAGATCCCTAAATACGTTCCACCTCAC 1504  
Db |||||  
859 IleTyrThrThrAspSerArgThrArgThrArgMetLeuLysTyrThrProGluHis 878  
Qy |||||  
1505 CTTTCTCCAGATAAGAGTGGCTTGA-----ACTCCCATTCAGGAGATGAGAAGA 1555  
Db |||||  
879 ThrTyrCysAsnAlaAlaPheTyrGlyProLeuCysSerPro-----Asn 893  
Qy |||||  
1556 ATGCTCGTGGTGGATCCGCTGCTCTCTTGAGACCATCTGCCAAT- 1603  
Db |||||  
894 ThrProPheCysGlyValGlnIleValAlaAsnSerAspThrGlyAsnGlyPheArgIle 913  
Qy |||||  
1604 -----CACACAGTAACTATTCGGGTAGATCTTTTCGGACGAGGAGAA 1645  
Db |||||  
914 AlaAlaThrGlyIleValGluGluIleAspValAsnIleGluIleValLysLeuLys 933  
Qy |||||  
1646 GTTCTCAACCTTTTCCAAACATTAATAAGATTTGTGGGATAACAAGCATGTTAAA- 1702  
Db |||||  
934 Leu---ValGlyPhePro-----TyrLysIlePheLysAsnThrAlaPheIleLysAsp 950  
Qy |||||  
1703 ATGCTTGTTCAGAACAAAATTTGTACCCAGTGAAGATGAGATGGTGAAGCACTCGG 1762  
Db |||||  
951 MetPheSerSerAlaMetGluValAlaArgPheGlyAlaGlnIleLysThrValSer 970  
Qy |||||  
1763 GGGAGCGGTGGGAGCTCATTCAGACTGCACCTTCAACAAATTTACAGACCCCAAAA 1822  
Db |||||  
971 GlyIleArgGlyGlu---IleLysArgAlaLeu-----SerLysProGluGly 985  
Qy |||||  
1823 -----TTGAGGATGCTATTCTGAAATACAATGTGGCATATTCTTAAGAAA 1867  
Db |||||  
986 HisTyrArgAlaAlaPheGluAspLysIleLeuMetSerAspIleValIleLeuArgSer 1005  
Qy |||||  
1868 TGGAGCTTTTACAGCTTTGATCGATTTCTGGATAGGTAAGTACTTTGAAGACGAGACTCAA 1927  
Db |||||  
1006 TrpTyrProValArgValLysLysPheTyrAsnProValThrSer----- 1020  
Qy |||||  
1928 CATTTATATCATGCTCATCTTGCCTGATATGTTGAAAATTGCATCTGTCTGCCAATATT 1987  
Db |||||  
1020 ----- 1020  
Qy |||||  
1988 TGCACCCAGCCCAATACCACTCTCTGAAACAGAGATGAATCATTCATCACAATGTGCGAG 2047  
Qy |||||

Best Local Similarity:	17.73%	Mismatches:	179
Query Match:	2.21%	Indels:	24
DB:	1	Gaps:	4
US-09-302-812-3 (1-4069) x SR40_YEAST (1-406)			
QY	281	CTGCACCCACAGACGCTCAGTGCAGTTCAGAGGTCCACCGTCCCTCGCAGCCTGGCTC	340
Db	14	LeuSerVallygGluLygGluGluGluLySerSerSerSerSerSerSerSerSerSer	33
QY	341	CCAGGGCAGCGGGCAGCAGCAGCAGCGCCACCTCGCTGCTTTTTCAAACAAAAGACT	400
Db	34	Ser	53
QY	401	ATTACCAAGTTGGATGACATAAAGGAATCAACAGACAGCGGAATCAGAAAGTTGGATAGT	460
Db	54	Ser	73
QY	461	AAAGAAAACAATACAGAAATAGAAATCCATGATGAGTTCTGTACAAAAGATAACTTT	520
Db	74	GluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	93
QY	521	TACCAACATAATGTAGAAAATTTAGTAAATGTTTCTCAGCTAAAGTCTGTATAAGTCACTC	580
Db	94	SerSerSerGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	113
QY	581	ACTGAAAAAGTACACAGTATTTTGAACACGATCAGACTGCACGCAATGTGTAAG---	637
Db	114	AspGluSerSerSerGluSerGluSerGluSerGluSerGluSerGluSerGluSer	133
QY	638	CAAAATGAAGGGAACACACGGAGAGCTTTTGGAAAGTCAACCTCAA-----	685
Db	134	AspAsnGluSerAlaLygGluThrLygAlaLygThrGluSerSerSerSerSerSer	153
QY	686	-----ACAGTAACCTGCTACACGACGAGTTTGTAAATGCT	721
Db	154	SerGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	173
QY	722	AACATTGATCGTCCACCTCAAAATGATGATCAGATGCACACAGATAGTGAAGAGAATAGA	781
Db	174	AspSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	193
QY	782	GACAATCAACAGTTTCTCACTGTAAAGTTGCGAAATGCAAGCGACGATACGGAAGAT	841
Db	194	AspSerGlnSer-----SerSerSerSerSerSerSerSerSerSerSerSer	206
QY	842	GAACACGCCAGAGCAACAAAGCCACCAGAGTGCAGCAGAGTCTTGCCATCTCGGGAA	901
Db	207	SerSerAspSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	226
QY	902	GACTGTGCAAGTTGTGACGAAGATGAGATAGACGTGTCGCAAGAGTCCATTGTTCAGAT	961
Db	227	Ser	246
QY	962	GTTGGCTCTCAGGATGTTGTACTGGGTCAAAAATATGACAACAAATTCATTACACAGAA	1021
Db	247	SerGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	262
QY	1022	AGTTGCTTAGGAATTTCTCTCTCCATTGTGAGAGGAAGTGAACCCGAATCCACCGATGGAT	1081
Db	263	SerThrSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	282
QY	1082	GTGGATAAATTTCTAAAATAGTTCTCAAGACTCAGAAGCAGATGAGGAGACAAGTCCAGGT	1141
Db	283	LeuGluThrLygGluAlaThrAlaAspGluSerLygAlaGluThrProAlaSerSer	302
QY	1142	TTTGATGAACAAGAGATGATGTTCTTCTCCAAACACGCAAAATTAACCTTCAAGTTCCAA	1201
Db	303	AsnGluSerThrProSerAlaSerSerSerSerSerSerSerSerSerSerSerSer	322
QY			



QY 397 KRDSKITDHF-----MRISSEDRKQCEVRHQTERKIPKYPNLPPEKWLGT 448  
 Db 616 SEGSMVGNFNSFDQNSYFSKASKRVPSPNIQI--SQNLPKGIA--LEPRR---T 667  
 QY 449 PIERMRMP--RCGIHLPSRPSASHTVTVRVLLRAGEVPKPPPTHY-----KDLWNKHV 503  
 Db 668 PFOSETKKELKHGTHOPAY-----PKKIPSPTRKHFFPAERTWNRQKI 711  
 QY 504 KMCSEQLYVDENGERTY--GSRWELIQTALINKETRP--QNLKDALILVNVAYSKKWD 561  
 Db 712 LPPLKED--YGRQDNLRHPSYSGRNI-----FYHEYINPYHNEKSOYIK-----SNPMD 760

RESULT 11  
 DPPP\_MOUSE STANDARD; PRT; 934 AA.

AC P97399; 070567;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Dentin sialophosphoprotein precursor (Dentin matrix protein-3) (DMP-3) [Contains: Dentin phosphoprotein (Dentin phosphoporphyrin) (DPP); Dentin sialoprotein (DSP)].  
 DE Dentin sialoprotein (DSP).  
 GS Mus musculus (Mouse).  
 ON Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=Swiss Webster; TISSUE=Molar;  
 RX MEDLINE=97150835; PubMed=8995371;  
 RA MacDougall M., Simmons D., Luan X., Nydegger J., Feng J.Q., Gu T.T.;  
 "Dentin phosphoprotein and dentin sialoprotein are cleavage products expressed from a single transcript coded by a gene on human chromosome 4." Dentin phosphoprotein DNA sequence determination.";  
 RT J. Biol. Chem. 272:835-842 (1997).  
 [2]  
 RN SEQUENCE FROM N.A. AND REVISIONS TO C-TERMINUS.  
 RC STRAIN=129/SvJ; TISSUE=Liver;  
 RX MEDLINE=98211969; PubMed=9545272;  
 RA Feng J.Q., Luan X., Wallace J., Jing D., Ohshima T., Kulkarni A.B., D'Souza R.N., Kozak C.A., MacDougall M.;  
 "Genomic organization, chromosomal mapping, and promoter analysis of the mouse dentin sialophosphoprotein (Dpp) gene, which codes for both dentin sialoprotein and dentin phosphoprotein.";  
 RT J. Biol. Chem. 273:9457-9464 (1998).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=129/SvJ; TISSUE=Liver;  
 RA Steir C., Butler S., Lin E., George A., Veis A.;  
 "From mouse to zebrafish-dentin matrix proteins genomic characterization.";  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=98055479; PubMed=9395101;  
 RA Ritchie H.H., Berry J.E., Somerman M.J., Hanks C.T., Bronckers A.L., Hutton D., Papegerakis P., Bernal A., Butler W.T.;  
 "Dentin sialoprotein (DSP) transcripts: developmentally-sustained expression in odontoblasts and transient expression in pre-odontoblasts.";  
 RT Eur. J. Oral Sci. 105:405-413 (1997).  
 [5]  
 RN TISSUE SPECIFICITY.  
 RP MEDLINE=21096982; PubMed=11175790;  
 RA Xiao S., Yu C., Chou X., Yuan W., Wang Y., Fu G., Qian M., Yang J., Shi Y., Hu L., Han B., Wang Z., Huang W., Liu J., Chen Z., Zhao G., Kong X.;  
 "Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP.";  
 RT Nat. Genet. 27:201-204 (2001).

CC -I- FUNCTION: DSP may be an important factor in dentinogenesis. DPP may bind high amount of calcium and facilitate initial mineralization of dentin matrix collagen as well as regulate the size and shape of the crystals.  
 CC -I- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -I- TISSUE SPECIFICITY: Expressed in teeth, mainly in odontoblasts and transiently in pre-odontoblasts. Found in the inner ear.  
 CC -I- PTM: DSP is glycosylated.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U67916; AAC12787.1; --; ALT\_SEQ.  
 CC EMBL; AJ002141; CRA05208.1; ALT\_SEQ.  
 CC EMBL; AF135799; AAD42781.1; ALT\_SEQ.  
 CC MGI; MGI:109172; Dppp.  
 CC Biomining; Extracellular matrix; Signal; Glycoprotein;  
 CC Phosphorylation; Sialic acid.  
 CC SIGNAL 1 17 POTENTIAL.  
 CC CHAIN 18 934 DENTIN SIALOPHOSPHOPROTEIN.  
 CC CHAIN 18 451 DENTIN SIALOPROTEIN.  
 CC CHAIN 452 934 DENTIN PHOSPHOPROTEIN.  
 CC DOMAIN 419 934 SER/ASP-RICH.  
 CC SITE 479 481 CELL ATTACHMENT SITE (POTENTIAL).  
 CC MOD\_RES 227 227 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 CC MOD\_RES 254 254 PHOSPHORYLATION (BY CK1) (POTENTIAL).  
 CC MOD\_RES 279 279 PHOSPHORYLATION (BY CK1) (POTENTIAL).  
 CC MOD\_RES 293 293 PHOSPHORYLATION (BY CK2) (BY SIMILARITY).  
 CC MOD\_RES 299 299 PHOSPHORYLATION (BY CK1) (BY SIMILARITY).  
 CC MOD\_RES 314 314 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 CC MOD\_RES 336 336 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 CC MOD\_RES 349 349 PHOSPHORYLATION (BY CK2) (POTENTIAL).  
 CC CARBOHYD 54 54 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 84 84 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 130 130 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 190 190 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 313 313 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC CARBOHYD 373 373 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC SEQUENCE 934 AA; 93901 MW; A618789D8A57249A CRC64;  
 Query Match 3.0%; Score 154; DB 1; Length 934;  
 Best Local Similarity 20.5%; Pred. No. 0.16;  
 Matches 80; Conservative 60; Mismatches 154; Indels 96; Gaps 14;  
 QY 34 RRRVLDPKDA-PVQPRVPPSSPACVSGRA-----GPHRGNAYSFVFKQITITWMD 84  
 Db 330 QDKEKLSPKDTRDAEGGIISQSEACPSGKSQDQGIETEGENKGN-----KSIIT--X 379  
 QY 85 TKGPKTAESKSNNTTRIDSMSSVQKDNFYHPKVKLENVQLNLDKSPTEKSSOYLN 144  
 Db 380 ESGKSGKSDSGHGHQGVDELKRNKSPKGSDEK-----QQTAKSAHSN 424  
 QY 145 QQQTASVCKWQNEGKHAQLALASEPPAGTPLPKQLSNANITGQSPHTTDDHSDTDEEDRN 204  
 Db 425 -----LGHRISSSSNSDGHGHD--SYEFDES 448  
 QY 205 QQPLTKPIKANTKPTVGDQCARNCCKSGSRQSKDCTCQQQEVLDVLPSPIDVGAED 264  
 Db 449 MQGDDP---KSSDESNGSDSDTNSANESGSRGDAYSYTSDESSDDDDSD--SHAGEDD 504  
 QY 265 IGTGPKNDKNTLGTQESSLGSDSPFEKESRPESPMVDNRSNCQDSEADDETSFVDEQD 324  
 Db 505 -----SSDDSGDSDSDSNGDG---DSESDKDESDSDHDNS--SDSEKSDSDSDSDSS 556  
 QY 325 DRSSQTANKLSSCQAREADQDLKRYLTKGSEVRLHFQEGENNAGTSDLNKPSGNSSS 384  
 Db 557 DSDSDSDSDSDSDSDSDSDS-----SDSD-----SSDSNSSSDSDSDSDSDSD 602



RT "the complete cDNA sequence of bovine coagulation factor V."; J. Biol. Chem. 267:2971-2978(1992). [2]

RP DISULFIDR BONDS.

RA MEDLINE-95034740; PubMed-7947716;

RX Xue J., Kafafis M., Silveira J.R., Kung C., Mann K.G.; "determination of the disulfide bridges in factor Va heavy chain."; Biochemistry 33:13109-13116(1994).

RT "FUNCTION: Coagulation factor V is a cofactor that participates with factor Xa to activate prothrombin to thrombin.

CC -!- SUBUNIT: Factor Va is composed of a heavy chain and a light chain, noncovalently bound. The interaction between the two chains is calcium-dependent.

CC -!- DOMAIN: Domain B contains 29.5 X 9 AA tandem repeats, and 2 X 14 AA repeats.

CC -!- PTM: Thrombin activates factor V proteolytically to the active cofactor, factor Va (formation of a heavy chain at the N-terminus and a light chain at the C-terminus).

CC -!- PTM: Sulfation is required for efficient thrombin cleavage and activation and for full procoagulant activity (By similarity).

CC -!- SIMILARITY: Contains 3 F5/8 type A domains.

CC -!- SIMILARITY: Contains 2 F5/8 type C domains.

CC -!- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

EMBL; M81440; AAA30512.1; -.

EMBL; M81441; AAA30513.1; -.

DR PIR; A42580; KFE05.

DR HSP; P12259; ICZT.

DR InterPro; IPR001117; Cu-oxidase.

DR InterPro; IPR008972; Cupredoxin.

DR InterPro; IPR000421; FA58 C.

DR InterPro; IPR008979; Gal\_Find\_like.

DR Pfam; PF00394; Cu-oxidase; 3.

DR Pfam; PF00754; F5\_F8\_type\_C; 2.

DR SMART; SM00231; FA58C; 2.

DR PROSITE; PS01285; FA58C.1; 2.

DR PROSITE; PS01286; FA58C.2; 2.

DR PROSITE; PS00022; FA58C.3; 2.

DR PROSITE; PS00079; MULTICOPPER OXIDASE1; 2.

KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen; Repeat.

FT SIGNAL 1 28

FT CHAIN 29 2211

FT CHAIN 29 741

FT PROPEP 742 1564

FT CHAIN 1565 2211

FT DOMAIN 30 327

FT DOMAIN 30 193

FT DOMAIN 203 327

FT DOMAIN 348 686

FT DOMAIN 348 525

FT DOMAIN 535 686

FT DOMAIN 696 1564

FT DOMAIN 1124 1151

FT REPEAT 1124 1137

FT REPEAT 1138 1151

FT DOMAIN 1188 1453

FT REPEAT 1188 1196

FT REPEAT 1197 1205

FT REPEAT 1206 1214

FT REPEAT 1215 1223

FT REPEAT 1224 1233

FT REPEAT 1241 1250

FT REPEAT 1251 1259

FT REPEAT 1260 1268

FT REPEAT 1269 1277

FT REPEAT 1278 1286

FT REPEAT 1287 1295

FT REPEAT 1296 1304

FT REPEAT 1305 1313

FT REPEAT 1314 1322

FT REPEAT 1323 1331

FT REPEAT 1332 1340

FT REPEAT 1341 1349

FT REPEAT 1350 1358

FT REPEAT 1359 1367

FT REPEAT 1368 1376

FT REPEAT 1377 1385

FT REPEAT 1386 1394

FT REPEAT 1395 1403

FT REPEAT 1404 1412

FT REPEAT 1413 1421

FT REPEAT 1422 1430

FT REPEAT 1431 1439

FT REPEAT 1440 1444

FT REPEAT 1445 1453

FT DOMAIN 1569 1890

FT DOMAIN 1569 1738

FT DOMAIN 1748 1890

FT DOMAIN 1894 2048

FT DOMAIN 2053 2208

FT SITE 741 742

FT SITE 1034 1035

FT SITE 1564 1565

FT DISULFID 167 193

FT DISULFID 248 329

FT DISULFID 499 525

FT DISULFID 607 688

FT DISULFID 1712 1738

FT DISULFID 1894 2048

FT DISULFID 2053 2208

FT MOD\_RES 697 697

FT MOD\_RES 701 701

FT MOD\_RES 730 730

FT MOD\_RES 1513 1513

FT MOD\_RES 1529 1529

FT MOD\_RES 1537 1537

FT MOD\_RES 1541 1541

FT CARBOHYD 225 225

FT CARBOHYD 239 239

FT CARBOHYD 297 297

FT CARBOHYD 382 382

FT CARBOHYD 460 460

FT CARBOHYD 553 553

FT CARBOHYD 587 587

FT CARBOHYD 745 745

FT CARBOHYD 756 756

FT CARBOHYD 774 774

FT CARBOHYD 780 780

FT CARBOHYD 902 902

FT CARBOHYD 952 952

FT CARBOHYD 964 964

FT CARBOHYD 1044 1044

FT CARBOHYD 1053 1053

FT CARBOHYD 1062 1062

FT CARBOHYD 1071 1071

FT CARBOHYD 1078 1078

FT CARBOHYD 1094 1094

FT CARBOHYD 1451 1451

FT CARBOHYD 1490 1490

FT CARBOHYD 1550 1550

FT CARBOHYD 1690 1690

FT CARBOHYD 1839 1839

FT REPEAT 1224 1233

FT REPEAT 1241 1250

FT REPEAT 1251 1259

FT REPEAT 1260 1268

FT REPEAT 1269 1277

FT REPEAT 1278 1286

FT REPEAT 1287 1295

FT REPEAT 1296 1304

FT REPEAT 1305 1313

FT REPEAT 1314 1322

FT REPEAT 1323 1331

FT REPEAT 1332 1340

FT REPEAT 1341 1349

FT REPEAT 1350 1358

FT REPEAT 1359 1367

FT REPEAT 1368 1376

FT REPEAT 1377 1385

FT REPEAT 1386 1394

FT REPEAT 1395 1403

FT REPEAT 1404 1412

FT REPEAT 1413 1421

FT REPEAT 1422 1430

FT REPEAT 1431 1439

FT REPEAT 1440 1444

FT REPEAT 1445 1453

FT DOMAIN 1569 1890

FT DOMAIN 1569 1738

FT DOMAIN 1748 1890

FT DOMAIN 1894 2048

FT DOMAIN 2053 2208

FT SITE 741 742

FT SITE 1034 1035

FT SITE 1564 1565

FT DISULFID 167 193

FT DISULFID 248 329

FT DISULFID 499 525

FT DISULFID 607 688

FT DISULFID 1712 1738

FT DISULFID 1894 2048

FT DISULFID 2053 2208

FT MOD\_RES 697 697

FT MOD\_RES 701 701

FT MOD\_RES 730 730

FT MOD\_RES 1513 1513

FT MOD\_RES 1529 1529

FT MOD\_RES 1537 1537

FT MOD\_RES 1541 1541

FT CARBOHYD 225 225

FT CARBOHYD 239 239

FT CARBOHYD 297 297

FT CARBOHYD 382 382

FT CARBOHYD 460 460

FT CARBOHYD 553 553

FT CARBOHYD 587 587

FT CARBOHYD 745 745

FT CARBOHYD 756 756

FT CARBOHYD 774 774

FT CARBOHYD 780 780

FT CARBOHYD 902 902

FT CARBOHYD 952 952

FT CARBOHYD 964 964

FT CARBOHYD 1044 1044

FT CARBOHYD 1053 1053

FT CARBOHYD 1062 1062

FT CARBOHYD 1071 1071

FT CARBOHYD 1078 1078

FT CARBOHYD 1094 1094

FT CARBOHYD 1451 1451

FT CARBOHYD 1490 1490

FT CARBOHYD 1550 1550

FT CARBOHYD 1690 1690

FT CARBOHYD 1839 1839

FT REPEAT 1224 1233

FT REPEAT 1241 1250

FT REPEAT 1251 1259

FT REPEAT 1260 1268

FT REPEAT 1269 1277

FT REPEAT 1278 1286

FT REPEAT 1287 1295

FT REPEAT 1296 1304

FT REPEAT 1305 1313

FT REPEAT 1314 1322

FT REPEAT 1323 1331

FT REPEAT



```

FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 NFIIPA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; CBBF908738667C45 CRC64;

Query Match 3.0%; Score 157.5; DB 1; Length 2211;
Best Local Similarity 18.7%; Pred. No. 0.32;
Matches 135; Conservative 102; Mismatches 295; Indels 189; Gaps 28;

Qy 65 HRGNATSFVFKQKITTMTMDTKGP----- 88
Db 641 HEDTLTLFPMQGESVNTVMDVGTWMLTMTNSPRSKKLRLFRDAKCIIRNDDSDSYEI 700
Qy 89 -----KTAESSEKNNTRIDSMSSVQKNFYPHKVEKL-----ENVQNLN 131
Db 701 YEPGSGTAMTTKKIHDSIEEDENDA---DSYQDELALILGLRSFRNSLSNQKDELNL 757
Qy 132 DKSPTEKSSQYLNOQQTASVCKWQEGKHABQLLASE-----PPAGTPIPLPKQ 178
Db 758 TALALEDKDSFIPPSANRSLDSNSSSRSHVSLAKNPAESLKTLLHLEAPAGSPLHA 817
Qy 179 LSNANIGOSPHTDHSDDHEDRDNOQF--LTPIKL-----ANTKPT-----VGDG 223
Db 818 GLDKNSALNPPMAEHS SPYSDPREDHPLSDVTGVSLLPFGTGFKRKPAPKHQRFQVGRG 877
Qy 224 QARNCKCSGR-QSVKCTCCQEQEVVDVLPESPLSDVGAEDIGTPKDNK-LTGQESS 281
Db 878 QAAKH-KESQTRFPAHKTRTLSDQNSSSRGPMWEDIPSDLLLLQQQDPKYLINGEWHL 936
Qy 282 LGDSPPPFKESPEPMDVDSNRNSCQ-----DSEADEETSPVF----- 320
Db 937 VSEKGSYELIQDANENKTVNKLNSPQDSRTWGENIPFKNSHGQSGHPTFLVTRRKL 996
Qy 321 -DEQDRSQTANKLSSCOAREADGLRKRYLTGSEVRLHPQPEGENNAGTSDLNKPS 379
Db 997 QDRQDRNRSLKEGLPLIRTRKKKEEPAYHVPLSPRSFH-PLRGEVNASFSRRHNHS 1055
Qy 380 -----GNSSSLNVECRSS--KQHGKRSKIDTFMWRISKSDRKRKEQCEVHRQTERKIP 432
Db 1056 LLLHASNETSLSIDLNQTFPPSMNLASLSPH-----DQTSNDTTSQTSSP 1103
Qy 433 KYIPNLPPEKKWLGTPTEERKMPRCGTHPLSLR-PSASHVTIV---RVDLIRAGEVPK 488
Db 1104 PDLYPTVSPHEHYQIFPIQD--SDPTHSTAPSNRSPDPTHSTAPSNRSPPTQSPQIPN 1161
Qy 489 PPTHYKDLWNKVMPCSEQLNYPVDENGERTAGRWELIQTALLNKTRPQNLKDA 548
Db 1162 -----YDLNRRAIPTDVSQIFP-----SLELVWQATSLDLSQP----- 1196
Qy 549 ILKYNVAYSKKWDFTALVDFMDKVLEEAQAHLVQSIPLDMVKIALCLPNICTQPI-PLL 607
Db 1197 -----SISPLDGMALSPDPGQ---ESLSPDLGQTSLS-PDLSQESLSPLD 1238
Qy 608 KQOMHVSVTMSQEQIASLANAFFECTPRRNAMKSEYSSYPDINFNRLF--EGSSRKP 665
Db 1239 GQTA-LSPDPSQESLSPDLGQTSLSDPD-----SQESLSPLDGTALSPPDGQESLSP 1290
Qy 666 E 666
Db 1291 D 1291

RESULT 10
ENAM MOUSE STANDARD; PRT; 1274 AA.
AC O55196;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Enamelin precursor.
GN ENAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Enamel epithelium;
RA MEDLINE=20514993; PubMed=11062988;
RX Hu C.-C., Simmer J.P., Bartlett J.D., Nanci A., Qian Q., Zhang C.,
RA Ryu O.H., Xue J., Fukae M., Uchida T., McDougall M.;
RT "Murine enamel: CDNA and derived protein sequences.";
RL Connect. Tissue Res. 39:47-61(1998).
CC -!- FUNCTION: Involved in the mineralization and structural
CC organization of enamel. Involved in the extension of enamel during
CC the secretory stage of dental enamel formation.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- TISSUE SPECIFICITY: Expressed in developing teeth.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U82698; AAB94312.1; -.
DR PIR: T37193; T37193.
DR MGI: MGI:1333772; Enam.
DR GO: GO:0005578; C:extracellular matrix; ISS.
DR GO: GO:0030345; F:structural constituent of tooth enamel; ISS.
DR GO: GO:0030282; P:bone mineralization; ISS.
DR GO: GO:0042476; P:odontogenesis; ISS.
KW Biomineralization; Extracellular matrix; Glycoprotein; Signal.
FT SIGNAL 1 38 POTENTIAL..
FT CHAIN 39 1274 ENAMELIN.
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1066 1066 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1274 AA; 140954 MW; F9BBD1CC9D327143 CRC64;

Query Match 3.0%; Score 156; DB 1; Length 1274;
Best Local Similarity 21.2%; Pred. No. 0.19;
Matches 127; Conservative 74; Mismatches 239; Indels 160; Gaps 32;

Qy 41 PKDAPQVRVPPSPACVSGPAGHRGNATSFVFKQKITTMTMDTKGPKTAESSEKNN 100
Db 242 PDDPPPEASTNSTVPDANATQSIPEGNDTSPI-----GNTGPGNAGNPP 287
Qy 101 TRIDSMSSVQKNFYPHKVEK-----LENVQLNL----- 131
Db 288 T-----VQGVFPFPKVNVSQGVKQSI PWRPQPNIVENYFVNPYSERQWQTG 339
Qy 132 DKSPTEKSSQYLNOQQTASVCKWQ---EGKHABQLLASEPPAGTPIPLKQLSNANIGOSP 188
Db 340 TQGPQNGQPGYRNQVQVRGP-QMNSFAWEGKQATR--PGNPTYGKP-PSPTSGVNYAGNP 395
Qy 189 HTDDHSDTDHEDRDNOQFL--TPIKLANTKPTVTDGQARSNCCKSCGSRQSKDCTGQQ 246
Db 396 -----VHFGRNLPQGNKPFVGANP---ASNKPFVFGANPASNKPFGVGNKPFVGN- 446
Qy 247 EVDVLPESPLSDVGAED-----IGTGPKNDNKLITG-----QESSLGSDSPPFKESEPE 295
Db 447 -----PASNKPFGVGNPASNKPFGVGNPASNKPFGVGNPASNKPFGVGN- 500
Qy 296 SP-MDVNSNRNSQDSEADEETSPVFDEQDRSRQTSQANK--LSSQAREAD-----GDL 346
Db 501 NFAANKPFVRNVGANKPFVGTNPSSNQPFRLSRNQSNKPFMRNQSNKPFVGTNVA5V 560
Qy 347 RRVLYTKGSEVRLHFOPEGENNAGTSDLNAPSGNSSSLNVECRSSKQHG----- 396
Db 561 GPKQVTVSHNMKT--QNPKEKSLGQKERTVTPTKDASN---PWRSAKQIGINNPNYLNPR 615

```

QY 385 LNVERGRSKQHGKRDSTKITDFMRISKSD 414  
 Db 603 SSDTCDSDSDSDSDSDSDSDSDSDSDSDSD 632

RESULT 12  
 PCLO RAT  
 ID PCLO RAT STANDARD; PRT; 5085 AA.  
 AC Q9JKS6; Q9JTL1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Piccolo protein (Multidomain presynaptic cytomatrix protein).  
 GN PCLO.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.  
 RX MEDLINE=20170257; PubMed=10707984;  
 RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,  
 RA Garner A.M., Kaempfi U., Kindler S., Gundelfinger E.D., Garner C.C.;  
 RA "Piccolo, a presynaptic zinc finger protein structurally related to  
 RT bassoon";  
 RT Neuron 25:203-214(2000).  
 RL [2]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;  
 RP VAL-4668; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND  
 RP ALA-4694.  
 RX MEDLINE=21181819; PubMed=11285225;  
 RA Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;  
 RA "An unusual C(2)-domain in the active-zone protein piccolo:  
 RT implications for Ca(2+) regulation of neurotransmitter release";  
 RL EMBO J. 20:1605-1619(2001).  
 CC -!- FUNCTION: May act as a scaffolding protein involved in the  
 CC organization of synaptic active zones and in synaptic vesicle  
 CC trafficking (By similarity).  
 CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic  
 CC junctions.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9JKS6-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9JKS6-2; Sequence=VSP 003930, VSP 003931;  
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
 CC phospholipids. Calcium binds with low affinity but with high  
 CC specificity and induces a large conformational change.  
 CC -!- SIMILARITY: Contains 2 C2 domains.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF138789; AAF07822.2; -;  
 CC DR EMBL; AF227534; AAF63196.1; -;  
 CC DR HSSP; P04410; 1A25.  
 CC GO; GO:0045202; C:synaptic junction; IDA.  
 CC GO; GO:0005509; F:calcium ion binding; IDA.  
 CC GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.  
 CC GO; GO:0005522; F:profilin binding; ISS.

GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.  
 GO; GO:0016080; P:synaptic vesicle targeting; NAS.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR001478; PDZ.  
 DR InterPro; IPR008899; Znf\_piccolo.  
 DR Pfam; PF00168; C2; 2.  
 DR Pfam; PF00595; PDZ; 1.  
 DR Pfam; PF05715; Zf\_piccolo; 2.  
 DR SMART; SM00239; C2; 2.  
 DR SMART; SM00228; PDZ; 1.  
 DR PROSITE; PS00499; C2 DOMAIN 1; 1.  
 DR PROSITE; PS00004; C2 DOMAIN 2; 2.  
 DR PROSITE; PS01006; PDZ; 1.  
 KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;  
 KW Repeat; Alternative splicing.  
 FT DOMAIN 372 491  
 FT 12 X 10 AA TANDEM APPROXIMATE REPEATS OF  
 FT P-A-K-P-Q-P-Q-P-X.  
 FT C4-TYPE (POTENTIAL).  
 FT C4-TYPE (POTENTIAL).  
 FT POLY-PRO.  
 FT PDZ.  
 FT C2 DOMAIN 1.  
 FT C2 DOMAIN 2.  
 FT TKPTN -> SKRRK (in isoform 2).  
 FT /FTID=VSP 003930.  
 FT Missing (in isoform 2).  
 FT /FTID=VSP 003931.  
 FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
 FT BINDING ACTIVITY.  
 FT D->A: COMPLETE LOSS OF CALCIUM-BINDING  
 FT AND CALCIUM-DEPENDENT PHOSPHOLIPID  
 FT BINDING ACTIVITY.  
 FT V->S: SMALL INCREASE IN AFFINITY FOR  
 FT CALCIUM.  
 FT VM->SS: 10-FOLD INCREASE IN AFFINITY FOR  
 FT CALCIUM.  
 FT M->S: INCREASED AFFINITY FOR CALCIUM.  
 FT VV->SS: 10-FOLD INCREASE IN AFFINITY FOR  
 FT CALCIUM.  
 FT QN->AA: MODERATE INCREASE IN AFFINITY FOR  
 FT CALCIUM.  
 FT A->S: NO EFFECT ON CALCIUM-BINDING  
 FT ACTIVITY.  
 SQ SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;

Query Match 3.0%; Score 153; DB 1; Length 5085;  
 Best Local Similarity 20.2%; Pred. No. 1.8;  
 Matches 122; Conservative 78; Mismatches 225; Indels 178; Gaps 26;

QY 75 KQRTITWMDTKGPKTAESKNNTRIDSMGSGVQKDNFYPHKVEKLVNPQLNLDKS 134  
 Db 1397 QOKTQTA-SETLDITISEEIKESQEKVSPKQDSEQ---FPSRKEHEK-PELVVDLS 1451  
 QY 135 PTERKSSQVLYNQQTASVCKWQNEGHQARLLASBPAGTLPKOLSNANIGQSPHTDDHS 194  
 Db 1452 PRASVYDSVED-----SSESENSPVVRKRRTSIGSS-SSDEYK 1489  
 QY 195 DTDHEEDRDNQFLTP--IKLANTKPT-----VGDGQAKSNCKCSG- 233  
 Db 1490 QEDSGSGSEEDFIRKQIIEISADESDSGSEDEFIRSQLKEISGVSGSKREAKGKGK 1549  
 QY 234 -----SRQSVKDCGCG-----QEEVDVLPESPLSPVGAED-----IGTG 268  
 Db 1550 GVAGKHRLTRKSTSTSFDDAGRRHSHWDEDETFDESPELKFRTKTSQSEELVVGAGG 1609  
 QY 269 PKNDNKLGTQESSLLGD-----SPFPEKESEPESPMDVDNSR--NSQCD 309  
 Db 1610 GLRRFKTIELNSTIADKYSSSESSQKKTLYDEEPELEMSLTDSPEDRSRSGSSSLHA 1669  
 QY 310 SEADETSPV----FDEQDRLSSQATANKLSSQCAEADGDLRKRYLTGSEVRLHFQEG 365  
 Db 1670 SSFTPTGTSVSVSLDESD--SSPHKKGESKQOKA-----RHRSHGPLL----- 1714

Query Match 3.0%; Score 152.5; DB 1; Length 406;  
 Best Local Similarity 20.3%; Pred. No. 0.066;  
 Matches 75; Conservative 58; Mismatches 163; Indels 73; Gaps 12;

366 ENNACTSDLNAPKSGNSLNVCESSKOHGKDSK--ITDHFMRISKSEDRKEQCEVR 423  
 1715 -----PTIEDSSEBELREELKEQKQRELEQOQKSSKSKKQKDEUR 1762

424 HORTERKIPKYPPLPPEKKWLGTPIDEMRKMPR-----CGIHLPLSPASHTVT 475  
 1763 AQR-RREPKTPPSNLSPED--ASPTIELRAAAEELHRSCSEYSPSTED----- 1813

476 VRVDLLRAGEVKKPFTTHYKDLWD-NKHVMPCCSQNLVYVDENGERTAGSRWBLIOTA 534  
 1814 -----PEGFESPEKIEIKVQVYKLPAAVSLYSPTDEQSVMOKEG-----VOKA 1857

535 LNKTRPNQNLKAILKYNVAISKWDFTAL--VDFWQKYLEAR-----AQHLYQSI 585  
 1858 -----LKSABEMEYEMQKPHKYKAFPAANERDEVFEKPELGGMLIEDIYIESL 1907

586 LPD 588  
 1908 VED 1910

## RESULT 13

SR40 YEAST STANDARD; PRT; 406 AA.  
 ID SR40 YEAST STANDARD; PRT; 406 AA.  
 AC P32583;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Suppressor protein SRP40.  
 GN SRP40 OR YKR092C OR YKR412A.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 26883 / FL1100;  
 RA Lalo D., Carles C., Sentenac A., Thuriaux P.;  
 RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94205265; PubMed=8154186;  
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,  
 RA Remacha M., Jimenez A., del Rey F.G., Revuelta J.L.;  
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces  
 cerevisiae chromosome XI contains the UBI2 and MPL1 genes and three  
 RT new open reading frames."  
 RL Yeast 9:1349-1354(1993).  
 CC -!- FUNCTION: Not known; weak suppressor of a mutant of the  
 CC subunit AC40 of DNA dependant RNA polymerase I and III.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L11275; AAA35091.1; -  
 CC EMBL; X73541; CAA51946.1; -  
 CC EMBL; Z28317; CAA82171.1; -  
 CC FIR; S38170; S38170.  
 CC GeneOnline; 140071; -  
 CC SGD; S0001800; SRP40.  
 CC GO; GO:0005730; C:nucleolus; IDA.  
 CC InterPro; IPR007718; SRP40\_C.  
 CC Pfam; PF05022; SRP40\_C; 1.  
 CC DOMAIN 25 314 ASP/SER-RICH.  
 CC FT 400 400 G -> N (IN REF. 1).  
 CC CONFLICT 406 AA; 41015 MW; 8EA007695AF4BA1D CRC64;  
 CC SEQUENCE

Query Match 3.0%; Score 152.5; DB 1; Length 406;  
 Best Local Similarity 20.3%; Pred. No. 0.066;  
 Matches 75; Conservative 58; Mismatches 163; Indels 73; Gaps 12;

75 KORTITTMOTGPKTAESKNNNTRIDSMSSVQKDNFYPHKVKEKLVNVPQNLNLDKS 134  
 17 KEKEI-----EKKSS 71

135 PTEKSSQVYNQOQTASVCWKQNEGHAEQLLASEPPAGTPLPKQLSNANIGOSPHTDHHS 194  
 72 DSESSSSSSSSSSSS-----SSDSSESSSESSSSSSSSSSSSSSSSSSSSSSSS 117

195 DTDHEEDR-----DNQOFLTPIKLANTKPTVGDGQARGNCKCGSRQSVKDTGCOQ 246  
 118 SESESEDETKKARESDNE- AKETKAKATEP---ESSSSSESSSSSSSSSSSSSSSES 173

247 E-----EVDVLPESPLSDVGAEDIGTGPKNDKLTQGESLGLSPPE 289  
 174 DSDSS 233

290 KESEPESPMDVNS-----RNSQDSEADEETSPVDFEQDDRSQTANKLSSCOAREADG 345  
 234 SDSDSS 291

346 LKRYLTGKSEVRLHFQPEGENNAGTSDLNAPKSGNSLNVCESSK-QHGKRDSPKIFD 404  
 292 ESKAETPAS-----SNSTPSASSSSSSANKLNIPAGTDEIKEGOR-----K 333

405 HPMRISKSR 413  
 334 HFSRVDRSK 342

## RESULT 14

PCLO\_CHICK STANDARD; PRT; 5120 AA.  
 ID PCLO CHICK STANDARD; PRT; 5120 AA.  
 AC Q9PU36;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Piccolo protein (Aczonin) (Fragment).  
 GN PCLO OR ACZ.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99439764; PubMed=10508862;  
 RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,  
 RA Killmann M.W.;  
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active  
 RT zones, shares homology regions with rim and bassoon and binds  
 RT profilin."  
 RL J. Cell Biol. 147:151-162(1999).  
 CC -!- FUNCTION: May act as a scaffolding protein involved in the  
 CC organization of synaptic active zones and in synaptic vesicle  
 CC trafficking (By similarity).  
 CC -!- SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of  
 CC synaptic junctions (By similarity).  
 CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and  
 CC phospholipids. Calcium binds with low affinity but with high  
 CC specificity and induces a large conformational change.  
 CC -!- SIMILARITY: Contains 2 C2 domains.  
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93139824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells.";
RC Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN [2]
RP REVISIONS.
RA Anderson S.K.;
RC Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: Component of a putative tumor-recognition complex.
CC Involved in the function of NK cells.
CC -1- SUBCELLULAR LOCATION: Membrane-anchored. Attached to the membrane
CC via its N-terminus.
CC -1- SIMILARITY: Contains 1 cyclophilin-like Pfiase domain.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L04288; AAA35734.2; -.
DR EMBL; AF184110; AAD56402.1; -.
DR PIR; A47328; A47328.
DR HSP; Q27450; IA33.
DR Genew; HGNC:7833; NKTR.
DR MIM; 161565; -.
DR GO; GO:0016018; F:cyclosporin A binding; TAS.
DR InterPro; IPR002130; CSA_Pfiase.
DR Pfam; PF00160; pro isomerase; 1.
DR PRINTS; PR00153; CERNPISMPASE.
DR PROSITE; PS00170; CSA_PFIASE_1; 1.
DR PROSITE; PS00072; CSA_PFIASE_2; 1.
DR KW Cyclosporin; Isomerase; Rotamase; Repeat; Membrane.
FT DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.
FT DOMAIN 219 240 ARG/LYS-RICH (BASIC).
FT DOMAIN 421 457 ARG/LYS-RICH (BASIC).
FT DOMAIN 970 1010 ARG/LYS-RICH (BASIC).
FT DOMAIN 194 244 ARG/SER-RICH.
FT DOMAIN 466 574 ARG/SER-RICH.
FT DOMAIN 664 814 ARG/SER-RICH.
FT DOMAIN 1311 1348 ARG-SER TANDEM REPEAT-RICH.
SQ SEQUENCE 1462 AA; 165676 MW; D98A1147763EF527 CRC64;
Alignment Scores:
Pred. No.: 0.00689 Length: 1462
Score: 159.00 Matches: 123
Percent Similarity: 34.34% Conservative: 93
Best Local Similarity: 19.55% Mismatches: 225
Query Match: 2.21% Indels: 188
DB: 1 Gaps: 23
US-09-302-912-3 (1-4069) x NKCR_HUMAN (1-1462)
Qy 392 CAAAGAGCACTATTACAGTTGGATGGACACTAAAGAAATCAAGACAGCGGAATCAGAAAGT 451
Db 675 GlnSerThrTyrSerLysTyrSerAspArgSerSerGluSerSerProArgSerArgSer 694
Qy 452 TTGATAGTAAA-----GAAACACACATCAAGATAGATCCATCATGAGT 499
Db 695 ArgSerSerArgSerArgSerTyrSerArgSerTyrThrArgSerArgSerLeuAlaSer 714
Qy 500 TCTGTACAAACAAAGAT-----AACTTTTACCAACATATAGTAAATTAGTAAAT 550

```

```

QY 1556 ATGCCTCGGTGGGATCGGCTGCTCTCTTGAGACCATCTGCCAATCACAGTAAC 1615
Db 1018 ----- 1018
QY 1616 ATTGGGTAGATCTTTTGGAGCAGGAGAAGTTCTTAACACTTTTCCACACATTATAA 1675
Db 1019 -----ProLeuGluPheGlyGluGluGlu 1027
QY 1676 GATTTGGGATTAACAGCATGTTAAATGCCCTTGTTCAGAACAAATTTGTACCCAGTG 1735
Db 1028 GluGluIleAspAspLysGlnValThrGlnGlnSerLysGluLysVal----- 1044
QY 1736 GAAGATGAGATGTCAGCAACTGCGGGAGCGGTGGAGCTCATTGAGCTGCACCT 1795
Db 1045 ----SerGluAsnAsnGlu----- 1049
QY 1796 CTCACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAATACAAATGTGGCA 1855
Db 1050 -----ThrIleLysAspAsnIleLeuLysThrGluLysSer 1061
QY 1856 -----TATTCTTAAGAAATGGACTTTACAGCTTTGATCGATTCTGGGATAAG 1903
Db 1062 SerGluGluAspLeuSerGlyLysHisAspThrValThrValSerSerAspLeuAspGln 1081
QY 1904 GTACTTGAAGACGACGAGCTCAACATTTATATCATGTCATCTTGCCTGATATGTTGAAA 1963
Db 1082 PheThrLysAspAspSerLysLeuSerIleSerProThrAlaLeuAsnThrGluGluAsn 1101
QY 1964 ATTGCACCTCTGTCGCAAAATATTGCAACCCCAATACCACTCTCTGAAACAGAAAGTG 2023
Db 1102 ValAla---CysLeuGlnAsnIle-----GlnHisVal 1111
QY 2024 ATCATCTCATCAATGTCGAGGACAGATGCGCAGTCTTTTACGATGCTTCTTCTTC 2083
Db 1112 GluGluSerValProAsnGlyValGluAspValLeuGlnThrAspAsnMetGluIle 1131
QY 2084 TGACATTTTCCAGCAGCAATGCTAAGATGAATCGAGTATCTAGT----- 2131
Db 1132 CysThr---ProAspArgSerSerProAlaLysValGluGluThrSerProLeuGlyAsn 1150
QY 2132 -----TACCACAGCATTAAC 2146
Db 1151 AlaArgLeuAspThrProAspIleAsn 1159

RESULT 14
SPCA HUMAN
AC P02549; Q15514; STANDARD; PRT; 2418 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Spectrin alpha chain, erythrocyte (Erythroid alpha-spectrin).
GN SPTA1 OR SPTA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP MEDLINE=90170949; PubMed=1689726;
RA Sahr K.E., Kotula L., Scarpa A.L., Coupal E., Leto T.L.,
RA Linnenbach A.J., Winkelman J.C., Speicher D.W., Marchesi V.T.,
RA Curtis P.J., Forget B.G.;
RT "The complete cDNA and polypeptide sequences of human erythroid
RT alpha-spectrin."
RL J. Biol. Chem. 265:4434-4443(1990).
[2]
RN
RP SEQUENCE OF 7-533 FROM N.A., AND VARIANTS EL2 PRO-260; PRO-261 AND
RP PRO-471.
RX MEDLINE=90009318; PubMed=2794061;
RA Sahr K.E., Tobe T., Scarpa A., Laughinghouse K., Marchesi S.L.,
RA Agre P., Linnenbach A.J., Marchesi V.T., Forget B.G.;

```

```

RT "Sequence and exon-intron organization of the DNA encoding the alpha
RT I domain of human spectrin. Application to the study of mutations
RT causing hereditary elliptocytosis."
RL J. Clin. Invest. 84:1243-1252(1989).
[3]
RN
RP SEQUENCE OF 7-601.
RX MEDLINE=84087888; PubMed=6654896;
RA Speicher D.W., Davis G., Marchesi V.T.;
RT "Structure of human erythrocyte spectrin. II. The sequence of the
RT alpha-I domain."
RL J. Biol. Chem. 258:14938-14947(1983).
[4]
RN
RP SEQUENCE OF 7-125.
RX MEDLINE=84087887; PubMed=6654895;
RA Speicher D.W., Davis G., Yurchenco P.D., Marchesi V.T.;
RT "Structure of human erythrocyte spectrin. I. Isolation of the alpha-I
RT domain and its cyanogen bromide peptides."
RL J. Biol. Chem. 258:14931-14937(1983).
[5]
RN
RP SEQUENCE OF 320-450 FROM N.A.
RX MEDLINE=86205962; PubMed=3458204;
RA Linnenbach A.J., Speicher D.W., Marchesi V.T., Forget B.G.;
RT "Cloning of a portion of the chromosomal gene for human erythrocyte
RT alpha-spectrin by using a synthetic gene fragment."
RL Proc. Natl. Acad. Sci. U.S.A. 83:2397-2401(1986).
[6]
RN
RP PARTIAL SEQUENCE.
RX MEDLINE=84295638; PubMed=6472478;
RA Speicher D.W., Marchesi V.T.;
RT "Erythrocyte spectrin is comprised of many homologous triple helical
RT segments."
RL Nature 311:177-180(1984).
[7]
RN
RP IDENTIFICATION OF PROBABLE FRAMESHIFT IN 2407-2418.
RA Gibson T.J.;
RL Unpublished observations (MAR-1995).
[8]
RN
RP REVIEW ON VARIANTS.
RX MEDLINE=97001215; PubMed=8844207;
RA Maillet P., Alloisio N., Morle L., Delaunay J.;
RT "Spectrin mutations in hereditary elliptocytosis and hereditary
RT spherocytosis."
RL Hum. Mutat. 8:97-107(1996).
[9]
RN
RP VARIANT EL2 SER-24.
RX MEDLINE=94289716; PubMed=8018926;
RA Farquet N., Devaux I., Boulanger L., Galand C., Boivin P.,
RA Lecomte M.-C., Dhermy D., Garbarz M.;
RT "Identification of three novel spectrin alpha I/74 mutations in
RT hereditary elliptocytosis: further support for a triple-stranded
RT folding unit model of the spectrin heterodimer contact site."
RL Blood 84:303-308(1994).
[10]
RN
RP VARIANTS EL2 CYS-28; HIS-28; LEU-28 AND SER-28.
RX MEDLINE=91358728; PubMed=1679439;
RA Coetzer T.L., Sahr K., Prchal J., Blacklock H., Peterson L., Koler R.,
RA Doyle J., Manaster J., Palek J.;
RT "Four different mutations in codon 28 of alpha spectrin are
RT associated with structurally and functionally abnormal spectrin alpha
RT I/74 in hereditary elliptocytosis."
RL J. Clin. Invest. 88:743-749(1991).
[11]
RN
RP VARIANT EL2 SER-28, AND VARIANT HPP ARG-48.
RX MEDLINE=91346849; PubMed=1878597;
RA Floyd P.B., Gallagher P.G., Valentino L.A., Davis M., Marchesi S.L.,
RA Forget B.G.;
RT "Heterogeneity of the molecular basis of hereditary
RT pyropoikilocytosis and hereditary elliptocytosis associated with
RT increased levels of the spectrin alpha I/74-kilodalton tryptic
RT peptide."
RL Blood 78:1364-1372(1991).
[12]
RN
RP VARIANT EL2 SER-45.

```

RL	Blood 74;828-832(1989).
RN	[20]
RP	VARIANT EL2 GENOVA TRP-34.
RR	MEDLINE=94250920; PubMed=8193371;
RX	Perrotta S., del Giudice E.M., Alloisio N., Sciarriatta G., Pinto L.,
RY	Delaunay J., Cutillo S., Lolascion A.;
RA	"Wild elliptocytosis associated with the alpha 34 Arg->Trp mutation
RT	in spectrin Genova (alpha I/74).";
RL	Blood 83;3346-3345(1994).
RN	[21]
RP	VARIANT EL2 ANASTASIA THR-45.
RR	MEDLINE=95290423; PubMed=7772539;
RX	Perrotta S., Tolascion A., de Angelis F., Pagano L., Colonna G.,
RY	Cutillo S., del Giudice E.M.;
RA	"Spectrin Anastasia (alpha I/78): a new spectrin variant (alpha 45
RT	Arg->Thr) with moderate elliptocytogenic potential.";
RL	Br. J. Haematol. 89:933-936(1995).
CC	-I- FUNCTION: Spectrin is the major constituent of the cytoskeletal
CC	network underlying the erythrocyte plasma membrane. It associates
CC	with band 4.1 and actin to form the cytoskeletal superstructure of
CC	the erythrocyte plasma membrane.
CC	-I- SUBUNIT: Composed of nonhomologous chains, alpha and beta, which
CC	aggregate side-to-side in an antiparallel fashion to form dimers,
CC	tetramers, and higher polymers.
CC	-I- DISEASE: Defects in SPHL are a cause of rhesus-unlinked
CC	elliptocytosis 2 (EL2) [MIM:130600, 182860, 166900].
CC	Elliptocytosis (also known as ovalocytosis) is a genetically
CC	heterogeneous, autosomal dominant hematologic disorder. It is

  

Alignment Scores:					
Pred. No.:	0.00787	Length:	2418		
Score:	159.00	Matches:	131		
Percent Similarity:	36.86%	Conservative:	99		
Best Local Similarity:	20.99%	Mismatches:	216		
Query Match:	2.21%	Indels:	178		
DB:	1	Gaps:	33		

  

US-09-302-812-3 (1-4069) x SPCA\_HUMAN (1-2418)

QY	401	ATTACCAAGTTGGATGGACACACTAAGAAGAAATCAAGACAGCGGAATCAGAAAGCTTGGTAGT	460
Db	594	LeuLysAsnTrpIleAsnLysLys--LysLeuAlaAspGlu-----AspTyr	610
QY	461	AAGAAAAACAACATACAGAAATAGATCATGTATGTTCTTCACAAAAGATAACTTT	520
Db	611	LysAsp-----IleGlnAsnLeuLysSerArgvalGlnLysGlnGlnVal	625
QY	521	TACCACATAATGTAGAAAAATTAGTAATGTCTTCAGCTA---AGTCTTGATAAGTCA	577
Db	626	PheGluLys-----GluLeuAlavalAsnLysThrGlnLeuGluAsnIleGlnLysThr	643
QY	578	CTCACTGAA-----AAAATACACAGTATTGAAACCAAGCATCAGATCGCAGCAATGTGT	631
Db	644	GlyGlnGlnMetIleGluGlyGlyHisTyralaserAspAsnValThrThrArgLysSer	663
QY	632	AAG-----TSGCAAATGAAGGGAACACACACGAGCAGCTTTTGGAAAGTGAA	679
Db	664	GluValalaserLeutr-----GluGluLeuLeuGluAlaThr	676
QY	680	CCTCAACAGTAACCTCGTACCAGAG-----CAGTTAGTAATGCTTAAC	724
Db	677	LysGlnLysGlyThrGlnLeuHisGlnAlaasnGlnGlnPheGluasn-----	694
QY	725	ATTGATCGGTCACTCAAAATGATGATCAGTCACACATAGTGAAGAGAAATAGAGAC	784
Db	695	-----AsnAlaGluAsp	698
QY	785	AATCAACAGTTCTCAACAAGTGTAAAGCTTGCAAATGCMAAGCAGACATCACGAAGATGAA	844
Db	699	LeuGlnArgTrpLeuGluAspValGluTrp-----GlnValThrSerGluAsp	714
QY	845	CACGCCAAGAAAGCCAAAGCCACAGAAAGTCAGACAGTCTTGCCATCTCTGGGAGAC	904
		:    :	









Qy	2099	-----CGAAATGCTAAGATGAATCGGAGTATTCTAGTTACCCA---	2137
Db	2084	ValThrIleLysGluThrGluThrAlaGluPhelYsAlaThrIleSerGlyPheProAla	2103
Qy	2138	-----GACATTAACTTCAATCGATTGTTGAGGACGTTTCATCAAGGAAA	2182
Db	2104	ProThrValLysTrpThrIleAsn---GluLysIleValGlu-----GluSerArgThr	2120
Qy	2183	CCGAGAGAACTTAAACGCTCTTCTGCTACTTTAGAAGAGTCACAGAGAAAAA	2236
Db	2121	IleThrThrIleLysThrGluAspValTyThrLeuLysIleSerAsnAlaLysIleGlu	2140
Qy	2237	CCTACTGGGTGGTGCACATTACAGACAGAGT-----CTTGAA	2275
Db	2141	GlnThrGlyThrValLysValThrAlaGlnAsnSerAlaGlyGlnAspSerLysGlnAla	2160
Qy	2276	GATTTTCCAGATGGGAAAGATGTGAAAAACCTTGCACAGATTGCATGTCACATTACGAA	2335
Db	2161	AspLeuLysValGluProAsnValLysAlaProLysPheLysSerGlnLeuThr-----	2178
Qy	2336	GGTACCATAGAGAAATGGCCAGGCATG-----CTACAGGTGCTTTTGCAAAAT	2386
Db	2179	AspLysValAlaAspGluGlyGluProLeuArgTrpAsnLeuLeuAsp-----	2195
Qy	2387	CGTTTGTGGAGGTGGTGTACCGAGTCAGGACTTGTGCCAAGAGAAATCGCTTTT	2446
Db	2196	-----GlyProSerProGly-----ThrGluValSerTrpLeu	2206
Qy	2447	ATCAATCCTGAGTTGATTATTTCACGGCTCTTCACTGAGGTGCTGATCACAAATGAA---	2503
Db	2207	LeuAsnGlyGlnProLeuThrLysSerAspThrValGlnValValAspHisGlyAspGly	2226
Qy	2504	TGCTTAATTATCAGAGGTACTGACAGTACAGTGAATACAGGC	2548
Db	2227	ThrTyHisValThrIleAlaGluAlaLysProGluMetSerGly	2241

Search completed: May 26, 2004, 16:52:39  
Job time : 235.042 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: May 26, 2004, 16:13:47 ; Search time 189,952 seconds

(without alignments)

13517.524 Million cell updates/sec

Title: US-09-302-812-3

Perfect score: 7208

Sequence: 1 ggcgcctgggaagtggagg.....agaaaaaaaaaaaaaaa 4069

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTO\_spool\_p/6333148/runat\_26052004\_150053\_5676/app\_query.fasta\_1.12437  
-DB=SPTREMBL\_25 -QWMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=6333148@cgn\_1\_1\_640@runat\_26052004\_150053\_5676 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25.\*

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rviris.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5190	72.0	976	4	Q9Y4W7

ID	Q9Y4W7	PRELIMINARY;	PRT;	976 AA.
AC	Q9Y4W7			
DT	01-NOV-1999 (TRENBLrel. 12, Created)			
DT	01-NOV-1999 (TRENBLrel. 12, Last sequence update)			
DT	01-JUN-2003 (TRENBLrel. 12, Last annotation update)			
DE	Poly(ADP-ribose) glycohydrolase.			
GN	HPARG			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]_SEQUENCE FROM N.A.			
RP	MEDLINE=99380098; PubMed=10449915;			
RX	Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;			
RT	"Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to human chromosome 10q11.23 and mouse chromosome 14B by in situ hybridization."			
RL	Cytogenet. Cell Genet. 85:269-270(1999).			
DR	EMBL; AF005043; AAB61614.1; --			
DR	Genbank; HGNC:8605; PARG.			
DR	GO; GO:0005737; Cytoplasm; TAS.			
DR	GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.			

#### ALIGNMENTS

RESULT 1

Q9Y4W7

Q9Y4W7 PRELIMINARY; PRT; 976 AA.

AC Q9Y4W7

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 12, Last annotation update)

DE Poly(ADP-ribose) glycohydrolase.

GN HPARG

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]\_SEQUENCE FROM N.A.

RP MEDLINE=99380098; PubMed=10449915;

RX Ame J.C., Apiau F., Jacobson E.L., Jacobson M.K.;

RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to human chromosome 10q11.23 and mouse chromosome 14B by in situ hybridization."

RL Cytogenet. Cell Genet. 85:269-270(1999).

DR EMBL; AF005043; AAB61614.1; --

DR Genbank; HGNC:8605; PARG.

DR GO; GO:0005737; Cytoplasm; TAS.

DR GO; GO:0004649; F:poly(ADP-ribose) glycohydrolase activity; TAS.

281	LeuIleArgGlnGlnSerCysLeuGlyAsnSerProPheGluIysGluSerGluPro	300
1067	GAATCACCGATGGATGGGATAATTCATAAAATAGTTGTCAAGCTCAGAAAGCAGATGAG	1126
301	GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu	320
1127	GAGACAAGTCCAGGTTTGTGATGAACAAGAAGATGGTAGTTCTCCCTCCAAAACAGCAAAATAAA	1186
321	GluThrSerProGlyPheAspGluGlnGluAspGlySerSerGlnThrAlaAsnLys	340
1187	CCTTCAAGGTTCCAAGCAPAGAGACGTGACATTTGAATTTAGGAAACGGTACTCTACTAAG	1246
341	ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys	360
1247	GGCGGTCAAGCTAGATTACATTTCCAAATTTGAAGGAGGAGAGACTCGCACCTGGGAATGAAT	1306
361	GlyGlyGluValArgLeuHisPheGlnPheGluGlyGluSerArgThrGlyMetAsn	380
1307	GATTAAATGCTAAACACTACCTCGGAAATATTCTAGCTCGAATGTAGAAATGCAGAAATCTCT	1366
381	AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer	400
1367	AAGCAACATGGAAAAAGGATTCATAAATCACAGATCATTTGATGAGACTGCCCAAGACA	1426
401	LysGlnHisGlyLysLysAspSerLysIleThrAspHisLeuMetArgLeuProLysAla	420
1427	GAGCACAGAGAAGAAACAGTGGGAAACCAAAACATCAAGAAACAGAAAGGAAGATCCCT	1486
421	GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro	440
1487	AAATACGTTTCCACTCCACCTTTCTCCAGATAAGAAGTGGCTGGAACTCCCATTTAGGAG	1546
441	LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu	460
1547	ATGAGAGAACTCCCTCGGTGGGATCCGGCTGCTCTCTTGAGACCATCTGCCAATCAC	1606
461	MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis	480
1607	ACAGTAACTATTCCGGGTAGATCTTTTTCGAGCAGGAGAAGTTCCTAAACCTTTTCCAACA	1666
481	ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr	500
1667	CATTATAAGATTTGTGGGATAACAAGATGTTAAATGCTCTCTCAGACAAATTTG	1726
501	HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu	520
1727	TACCCAGTGAAGATGAGATGGTGAGCGAACTCGGGGAGCCGGTGGGAGCTCATTTACAG	1786
521	TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln	540
1787	ACTGCACCTTCTCAACAAATTTTACACGACCCCAAACTTGAAGGATGCTATTCTGAATAC	1846
541	ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr	560
1847	AATGTGGCATATTCTAGAATGGCACTTACAGCTTTGATCGATTTCTGGGATAAGGTA	1906
561	AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal	580
1907	CTTCAAGAAGCAGAGGCTCAACATTTATATATATATATATATATATATATATATATATAT	1966
581	LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle	600
1967	GCATCTGTCTGCCAAATATTTTGCACCCAGCAATACACCTCTCGAAACAGAAATGAAT	2026
601	AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMetAsn	620
2027	CATTCCATCAAAATGTCGAGGAACAGATGCCAGTCTTTTAGCTTAATGCTTTCTTCGCG	2086
621	HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys	640
2087	ACATTTCCACGACGAATGCTTAAGATGAATTCGAGATTTCTTAGTTACCCAGACATTAAC	2146
641	ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn	660

QY 2147 TTCAATCGATTGTTGAGGACGCTTCATCAAGAAACCGAGAACTTAAACGCTCTTC 2206  
 DB 661 PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe 680  
 QY 2207 TGCTACTTTAGAGAGTACAGAGAAACCTACTGGGTGGTGCATTTTCAAGACAG 2266  
 DB 681 CysTyrPheArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln 700  
 QY 2267 AGCTTGAAGATTTCCAGAAAGTGAAGATGAAACCCCTTGACACCAATTCATGTC 2326  
 DB 701 SerLeuGluAspPheProGluTyrGluArgCysGluLysProLeuThrArgLeuHisVal 720  
 QY 2327 ACTTACGAAGTACATAGAGAAATGCGCAAGGATGCTACAGGTGGATTTTGCAT 2386  
 DB 721 ThrTyrGluGlyThrLeuGluLysGlyGlyMetLeuGlnValAspPheAlaAsn 740  
 QY 2387 CGTTTGTGGAGTGGTGTAAACAGTGCAGGACTTGTGCAAGAAGAAATCCGCTTTTA 2446  
 DB 741 ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluLysLeuArgPheLeu 760  
 QY 2447 ATCAATCCTGAGTGAATTTTACGCTCTTCACTGAGTGTGATCAATGATATGT 2506  
 DB 761 IleAsnProGluLeuLeuIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys 780  
 QY 2507 CTAATTATCAGTACTGAGCAGTACAGTGAATACACAGGCTATGCTGAGACATATCGT 2566  
 DB 781 LeuLeuIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg 800  
 QY 2567 TGTTCGCGAGCCAGAAAGTGGAGTGAAGGAGCACTCGAGCGCGCTGCCTGAG 2626  
 DB 801 TrpSerArgSerHisGluAspGlySerGluArgAspAspCysGluArgCysThrGlu 820  
 QY 2627 ATGCTGCCATCGATGCTCTTCACTCAGACGCTACCTCGATCAGTTTGGCTGAGAAA 2686  
 DB 821 IleValAlaIleAspAlaLeuHisPheArgTyrLeuAspGlnPheValProGluLys 840  
 QY 2687 ATGAGCGCGAGTGAACAGGCTTACTGCGATTTCTCGCTGCGAGTTCTTCAGAG 2746  
 DB 841 MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu 860  
 QY 2747 AATCTTCTCAGTGGCCAGGAAACTGGGGCTGTGCTGCTTGGGGGTGATGCCAGG 2806  
 DB 861 AsnLeuSerAlaValAlaThrGlyAsnTyrGlyCysGlyAlaPheGlyGlyAspAlaArg 880  
 QY 2807 TTAAGACCTTAATACAGATATTGGCAGCTGCTGAGCTGAGCAGATGGTTTATTTC 2866  
 DB 881 LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyrPhe 900  
 QY 2867 ACCTTTGGGACTCAGAAATGATGAGACATTTACAGCATGACATTTTCTTACTGAA 2926  
 DB 901 ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisLeuPheLeuThrGlu 920  
 QY 2927 AGGAACTCACTGTGGAGATGTGTAAAGCTGTGTACGATCACTACAATGAAGATGC 2986  
 DB 921 ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys 940  
 QY 2987 AGAACTGTTCACCCCTCGACACGACATCAGCTTTATCCATTCATATACCATGCTGC 3046  
 DB 941 ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal 960  
 QY 3047 GAGTCTGTGAGACGCTGACCATTCAGGCAAGAGGACAGGACC 3094  
 DB 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976

## RESULT 2

Q86W56 PRELIMINARY; PRT; 976 AA.  
 AC Q86W56  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Similar to poly (ADP-ribose) glycohydrolase.

OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strausberg R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC050560; AAH50560.1; -  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE. 976 AA; 111109 MW; D646353C6D0180E CRC64;

## Alignment Scores:

Pred. No.: 0 Length: 976  
 Score: 5113.00 Matches: 963  
 Percent Similarity: 98.98% Conservative: 3  
 Best Local Similarity: 98.67% Mismatches: 10  
 Query Match: 70.94% Indels: 0  
 DB: 4 Gaps: 0

US-09-302-812-3 (1-4069) x Q86W56 (1-976)

QY 167 ATCAATCGCGCCCGCTGTGAACCTGCACCAAGCGACCGCTGGGCGCGCTACA 226  
 DB 1 MetAsnAlaGlyProGlyCysGluProCysThrLysArgProArgTyrGlyAlaAlaThr 20  
 QY 227 ACTTCCCGCGCTCTTCGAGACCGCGAGCTTTCGAGCAGGAGCGCGCTCTCGAC 286  
 DB 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgValLeuAsp 40  
 QY 287 CCCAAGCGCTCACTGAGTTCAGGTCAGGTCACCGCTCTGCGCAGCTGCTCCAGGG 346  
 DB 41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60  
 QY 347 CAGCGCGGACAGCAGAGCGCGCCACTGCTGTTTTCACAAAGAGATATTACC 406  
 DB 61 ArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80  
 QY 407 AGTTGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGGATAGTAAGAA 466  
 DB 81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100  
 QY 467 AACAAATACAGATAGATCCATGATGAGTTCGTGTAACAAAGATATCTTTACCA 526  
 DB 101 AsnAsnAsnThrArgIleGluSerMetSerValGlnLysAspAsnPheTyrGln 120  
 QY 527 CATAAATGTAGAAAAATTAGTAAATGTTCTCAGCTAAGTCTTGATAGTCACTCACTGAA 586  
 DB 121 HisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThrGlu 140  
 QY 587 AAAAGTACAGATATTGAAACCAAGCATCAGACTGACCAATGTGTAAAGTGGCAAAATGAA 646  
 DB 141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysTyrGlnAsnGlu 160  
 QY 647 GGGAAACACACGAGCAGCTTTGGAAAGTGAACCTCAACAGTAACTCCCTGGTACCAG 706  
 DB 161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180  
 QY 707 CAGTTTAGTAATCTAACATGATCGTCACTCAAAATGATCATCACAGTGCACAGAT 766  
 DB 181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspAspHisSerAspThrAsp 200  
 QY 767 AGTGAAGAGATAGACAAATCAACAGTTTCTCACAACCTGTAAAGCTTGCAATGCAAG 826  
 DB 201 SerGluGluAsnArgAspAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220  
 QY 827 CAGACTACGGAAGATGAACACGCCAGAGAGCCAAAGCCACCAAGAGTGCACAGTCT 886  
 DB 221 GlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240

QY	887	TGCCATCTCTGGGAGACGTGCTCAAGTTGTCACGAGATGAGATAGACGTGTCACAAAG	946	QY	1967	GCACTCTCTGCTCCAAATATTTTGACCCAGCCAGCAATACCACTCTCTGAAACAGAGATGAT	2026
Db	241	CysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluLeuAspValValProGlu	260	Db	601	AlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuGlnLysMetAsn	620
QY	947	AGTCATTTCTCAGATGTTGGCTCTCAGGATGTTGGTACTCGGTCACAAAATGACACAAA	1006	QY	2027	CATTCCATCACAATGTCGAGAACAGATTGCGAGCTTTTAGCTTAATGCTTTCTCTGTC	2086
Db	261	SerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsnLys	280	Db	621	HisSerIleThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys	640
QY	1007	TTGATTAGACAAGAAGTTGCCCTAGGAATTTCTCTCCATTTGAGAGAGAAAGTGAACCC	1066	QY	2087	ACATTTCACACGAGAAATGCTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG	2146
Db	281	LeuThrArgGlnGluSerCysLeuGlyAsnSerProPheGluLeuSerGluPro	300	Db	641	ThrPheProArgAsnAlaLysMetLysSerGluLysSerSerLysProAspIleAsn	660
QY	1067	GAATCACCGATGATGCTGATAATTTCTAAATAGTTGTCAAGACTCAAGACAGATGAG	1126	QY	2147	TTCAATTCATGTTGAGGAGCTTCATCAAGAAACCGGAGAACTTAAACCGCTCTTC	2206
Db	301	GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu	320	Db	661	PheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeuPhe	680
QY	1127	GAGACAAGTCCAGGTTTTCATGAACAAGAGATGAGTGTCTCTCCAAACAGCAATAAA	1186	QY	2207	TGCTACTTTAGAGAGTCAAGAGAAAAAACCCTACTGGGTGGTGCATTTACAGACAG	2266
Db	321	GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys	340	Db	681	CysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln	700
QY	1187	CCTTCAAGGTTCCACAGACAGCGCTGACATTGAATTTAGGAACGGTACTCTACTAAG	1246	QY	2267	AGTCTTGAAGATTTTCCAGAAATGGAAAGATGTGAAAAACCCCTTCACACGATTGCTATC	2326
Db	341	ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys	360	Db	701	SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal	720
QY	1247	GGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGAGAGAGAGTCCGCACTGGAATGAT	1306	QY	2327	ACTTACGAAGGTACCATAGAAGAAATGCCAAGCATGCTACAGGTGGATTGTCAAAT	2386
Db	361	GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn	380	Db	721	ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn	740
QY	1307	GATTTAAATGCTAACTACTCGAAATATTTCTAGCCTGAATGTAGATGCGAATTTCT	1366	QY	2387	CGTTTTGTTGGAGTGTGTAAACAGTGCAGACACTGTGTCAGAGAGAAATCGCTTTT	2446
Db	381	AspLeuAsnAlaLysLeuProGlyAsnIleSerSerLeuAsnValGluCysArgAsnSer	400	Db	741	ArgPheValGlyGlyGlyValThrSerAlaGlyLeuValGlnGluGluLeuArgPheLeu	760
QY	1367	RAGCAATGGAAGAAAGATTTCTAAATACACAGATCATTTGATGAGCTGCCCAAGCA	1426	QY	2447	ATCAATCTCGAGTTCATTTATTTACGGCTCTTCACTGAGGTGCTGCTGATCACAATGAT	2506
Db	401	LysGlnHisGlyLysAspSerLysIleThrAspHisPheMetArgLeuProLysAla	420	Db	761	IleAsnProGluLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys	780
QY	1427	GAGGACAGAAAGAAAGACAGTGGGAAACCAACATCAAGAACAGAAAGGAGATCCCT	1486	QY	2507	CTAATTCACAGGTACTGACAGTACAGTGAATACACAGGCTATGCTGAGACATATGCT	2566
Db	421	GluAspArgArgGlyGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro	440	Db	781	LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg	800
QY	1487	AAATACGTTCCACCTTCTCCAGATAGAAGTGGCTTGGAACTCCCATTTGAGGAG	1546	QY	2567	TGCTCCCGAGCCACAGAGATGGAGTGAAGAGGAGCATCGCAGCGCGCTGCTACGAG	2626
Db	441	LysTyrValProProHisLeuSerProAspLysLysTyrLeuGlyThrProIleGluGlu	460	Db	801	TrpSerArgSerHisGluAspGlySerGluArgAspAspTrpGlnArgCysThrGlu	820
QY	1547	ATGAGAGATGCTCGGTGGATCCGGCTGCTCTCTGAGACCATCTGCCATCAC	1606	QY	2627	ATCGTTGTCATCGATGCTCTTCACTCAGAGCTACCTCGATCAGTTGTCCTCAGAAA	2686
Db	461	MetArgArgMetProArgCysGlyIleArgLeuProLeuLeuArgProSerAlaAsnHis	480	Db	821	IleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGluLys	840
QY	1607	ACAGTAACCTATTTCGGGTAGATCTTTTGGCAGCAGAGAGATTCCTAAACCTTTTCAACA	1666	QY	2687	ATGAGACGCGAGCTGAAACAAGCTTACTGTGATTTCTCCGTCCTGAGATTCTTCAGAG	2746
Db	481	ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr	500	Db	841	MetArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSerGlu	860
QY	1667	CATTATAAGATTTGGGATACAAAGCATGTTAAATGCTTGTTCAGAACAAATTTG	1726	QY	2747	TAATCTTCTGAGTGGCCACAGGAACTGGGCTGTGTGCTCTTGGGGGTGATGCCAGG	2806
Db	501	HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu	520	Db	861	AsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAlaArg	880
QY	1727	TACCCAGTCGAGATGATGTCAGCAACTCGCGGAGCGGTGGGAGCTCATTTAG	1786	QY	2807	TTAAAGGCTTAAATACAGATATTGGCAGCTGCTGAGCTGAGCAGATGTGTTTATTTC	2866
Db	521	TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln	540	Db	881	LeuLysAlaLeuIleGlnIleLeuAlaAlaAlaAlaGluArgAspValValTyrPhe	900
QY	1787	ACTGCACTCTCTCAACAAATTTACACACCCCAAACTTGAAGGATGCTATTCTGAATAC	1846	QY	2867	ACCTTTGGGACTCAGAATTGATGAGACATTTACAGCATGCCATTTCTTACTGAA	2926
Db	541	ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr	560	Db	901	ThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisIlePheLeuThrGlu	920
QY	1847	AATGTCGATTTCTAAGAAATGGGCTTTACAGCTTTGATGATTTCTGGGATAGGTA	1906	QY	2927	AGGAACCTCACCTGTTGGAGATGTGTATAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2986
Db	561	AsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal	580	Db	921	ArgLysLeuThrValGlyAspValTyrLysLeuLeuLeuArgTyrTyrAsnGluGluCys	940
QY	1907	CTTGAAGACGACAGCTCAACATTTATATCAGTCCATCTTGCCTCATATGTTGAAAT	1966	QY	2987	AGAAACTGTTCCACCCCTGGACAGACATCAAGCTTTATCCATTCATATACCATGCTGTC	3046
Db	581	LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle	600	Db	941	ArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAlaVal	960
				QY	3047	GAGTCTGTGCAGAGACCGCTGACCATTCAGCGGCAAGAGACAGGAGACC	3094

Db 961 GluSerCysAlaGluThrAlaAspHisSerGlyGlnArgThrGlyThr 976  
 |||||

RESULT 3  
 Q72742 ID Q72742 PRELIMINARY; PRT; 976 AA.  
 AC Q72742; MEDLINE=22388257; PubMed=12477932;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullan S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Skalska J., Schnerch D.E., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Strausberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC052966; AAH52966.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 976 AA; 111090 MW; 47C9EE1826C4A74A CRC64;  
 Alignment Scores:  
 Pred. No.: 0 Length: 976  
 Score: 5108.00 Matches: 962  
 Percent Similarity: 98.87% Conservative: 3  
 Best Local Similarity: 98.57% Mismatches: 11  
 Query Match: 70.87% Indels: 0  
 DB: 4 Gaps: 0  
 US-09-302-812-3 (1-4069) x Q72742 (1-976)  
 QY 167 ATCAATCGCGCGCGCGCTGTGAACCTCGACCAAGCGACCCGCTGGGCGCGCTACA 226  
 Db 1 MetAsnAlaGlyProGlyCysGluProCysThrIysArgProArgTTPGlyAlaAlaThr 20  
 QY 227 ACTTCGCGCGCTGTTCGGAGCCCGCGAGCTTCCGAGCAGGCGAGCGCGGCTCTGAC 286  
 Db 21 ThrSerProAlaAlaSerAspAlaArgSerPheProSerArgGlnArgAvalLeuAsp 40  
 QY 287 CCCAAGCAGCTCAGCGTGCAGTTCAGGCTCCACCGCTCCGCGCGCTGCTGCCAGGG 346  
 Db 41 ProLysAspAlaHisValGlnPheArgValProProSerSerProAlaCysValProGly 60  
 QY 347 CAGCGCGGACAGCAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 406  
 Db 61 ArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIleThr 80  
 QY 407 AGTTGGATGCACTAAAGAGATCAAGACGCGGATCAGAAAGTTTGGATAGTAAGAA 466  
 Db 81 SerTrpMetAspThrLysGlyIleLysThrAlaGluSerGluSerLeuAspSerLysGlu 100  
 QY 467 AACCAACAATACAGAATAGATCCATGATGAGTTCGTACAAAAGATAACTTTTACAA 526  
 Db 101 AsnAsnAsnThrArgIleGluSerMetMetSerSerValGlnLysAspAsnPheTyrGln 120  
 QY 527 CATATGTGAGAAAAATAGTAATGTTTCTCAGCTAAGTCTTGTATAGTCACCTCAGTAA 586  
 Db 121 HisAsnValGluLysLeuGluAsnValSerGlnLeuSerLeuAspLysSerProThrGlu 140  
 QY 587 AAAAGTACACAGTATTTGACACAGCATCAGCTGCAGCATGTGTAGTGGCAAAATGAA 646  
 Db 141 LysSerThrGlnTyrLeuAsnGlnHisGlnThrAlaAlaMetCysLysIleGlnAsnGlu 160  
 QY 647 GGGAAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACACAGTAAACCTGGTACAGAG 706  
 Db 161 GlyLysHisThrGluGlnLeuLeuGluSerGluProGlnThrValThrLeuValProGlu 180  
 QY 707 CAGTTTGTAGTAATGCTAACATGATTCGGTCACTCAAAATGATGATCATCATGACAGAT 766  
 Db 181 GlnPheSerAsnAlaAsnIleAspArgSerProGlnAsnAspHisSerAspThrAsp 200  
 QY 767 AGTGAAGAGATAGACACATCAACAGTCTTCTCACAACTGTAAAGCTTGCATAATGCAAG 826  
 Db 201 SerGluGluAsnArgAsnGlnGlnPheLeuThrThrValLysLeuAlaAsnAlaLys 220  
 QY 827 CAGACTACGGAAGATGAACACACGCGCAGAGAGCCAAAGCCACAGAGTGCACCAAGTCT 886  
 Db 221 GlnThrThrGluAspGluGlnAlaArgGluAlaLysSerHisGlnLysCysSerLysSer 240  
 QY 887 TGCATCTCTGGGAGACCTGTGCAAGTTGTTCAGCAGATGAGATAGACGTGGTGCCTAAG 946  
 Db 241 CysAspProGlyGluAspCysAlaSerCysGlnGlnAspGluIleAspValValProGlu 260  
 QY 947 AGTCCATCTGTGAGATCTGCTCTGAGGATGTTGTTACTGGGTCAAAAATGCAACAAA 1006  
 Db 261 SerProLeuSerAspValGlySerGluAspValGlyThrGlyProLysAsnAspAsnLys 280  
 QY 1007 TTGATTAGACAAAGATTTGCCCTAGGAAATTTCTCCCTCAATTTGAGAGAGAAAGTGAACCC 1066  
 Db 281 LeuThrArgGlnGluSerCysLeuGluAsnSerProProPheGluLysGluSerGluPro 300  
 QY 1067 GAATCACCGATGATGTGATTAATTTCTAAATATGTTGTCAACTCAGACTCAGAGCAGATGAG 1126  
 Db 301 GluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAspGlu 320  
 QY 1127 GAGACAAGTCCAGCTTTGATGAACAAGAGATGGTAGTTCTCCCTCCCAACAGCAAAATAAA 1186  
 Db 321 GluThrSerProGlyPheAspGluGlnGluAspGlySerSerSerGlnThrAlaAsnLys 340  
 QY 1187 CCTTCAAGGTTCCAGCAAGACGCTGACATTTAGAAATTTAGGAAACGGTACTCTACTAAG 1246  
 Db 341 ProSerArgPheGlnAlaArgAspAlaAspIleGluPheArgLysArgTyrSerThrLys 360  
 QY 1247 GCGGCTGAGTTAGATTACATTTCCATTTGAAGGAGGAGAGTCGCTCGAATGAAT 1306  
 Db 361 GlyGlyGluValArgLeuHisPheGlnPheGluGlyGlyGluSerArgThrGlyMetAsn 380  
 QY 1307 GATTTAATCTCAAACTTACTCGAAATATTTCTAGCTGTAATGTAGAATGCAGAAATTTCT 1366  
 Db 381 AspLeuAsnAlaLysLeuProGlyLysIleSerSerLeuAsnValGluCysArgAsnSer 400  
 QY 1367 AAGCAACATGGAAAAAGGATTTCAAAATCAGAGATCATTTGTATGAGACTTGCCTCAAGCA 1426  
 Db 401 LysGlnHisGlyLysAspSerLysIleThrAspHisPheMetArgLeuProLysAla 420  
 QY 1427 GAGACAGAAGAAAAAGACAGTGGGAAACCAACATCAAGAAACAGAAAGAGAGATCCCT 1486  
 Db 421 GluAspArgArgLysGluGlnTrpGluThrLysHisGlnArgThrGluArgLysIlePro 440



Qy	1487	AAATACGTTCCCACTCACTTTCTCCAGATAAGAGTGGCTTGGAACTCCCATTTGAGGAG	1546
Db	441	LysTyrValProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGluGlu	460
Qy	1547	ATCAGAGAAATGCTCGTGGTGGATCCGGCTCCCTCTTTGAGACCATCTGCGCAATCAC	1606
Db	461	MetArgArgMetProArgCysGlyIleArgLeuProIleuArgProSerAlaasnHis	480
Qy	1607	ACAGTAACTATTGGGTAGATCTTTTCGACGACGAGAGTCTCTAAACCTTTTCAACA	1666
Db	481	ThrValThrIleArgValAspLeuLeuArgAlaGlyGluValProLysProPheProThr	500
Qy	1667	CATTATAAGATTGTCGGATAACAACATGTTAAATGCCTTGTTCAGAACAAAAATTTC	1726
Db	501	HisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsnLeu	520
Qy	1727	TACCCAGTGGAAATGAGATGGTAGCGAACTCGGGGAGCCGGTGGGAGCTCATTCAG	1786
Db	521	TyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIleGln	540
Qy	1787	ACTGCATCTTCAACAAATTTCACGACCCCAAACTTTGAAGGATGCTATTCTGAATAC	1846
Db	541	ThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLysTyr	560
Qy	1847	AATGTGCATATTCTAAGAATGGCACTTTACAGCTTTGATCGATTTCTCGGATAAGGTA	1906
Db	561	AsnValAlaItyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLysVal	580
Qy	1907	CTTGAAGAAGCAGAAGCTCAACATTTATATCAGTCCATCTGTGCTGATATGGTGAATAAT	1966
Db	581	LeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLysIle	600
Qy	1967	GCACCTGCTGCCAAATATTTCACCCAGCCCAATACCATCTCTGAAACAGAAGATGAAT	2026
Db	601	AlaLeuCysLeuProAsnIleCysThrGlnProIleProIleuLeuLysGlnLysMetAsn	620
Qy	2027	CATTCCATCAATGTCCGAGAACAGATTGCCAGTCTTTTAGCTAATGCTTTCTTCTGC	2086
Db	621	HisSerIleThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePheCys	640
Qy	2087	ACATTTCCACGACGAATGCTAAGATGAATCGAGTATTCTAGTTACCCAGACATTAAC	2146
Db	641	ThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIleAsn	660
Qy	2147	TTCAATCGATTGTTGAGGACGTTTCATCAAGGAAACCGGAGAACTTAAACCGCTCTTC	2206
Db	661	PheAsnArgLeuPheGluItyrArgSerSerArgLysProGluLysLeuLysThrLeuPhe	680
Qy	2207	TGCTACTTTAGAAGCTCACAGAGAAAAACCTACTGGTTGGTGACATTTACAGACAG	2266
Db	681	CysTyrPheArgValThrGluLysLysProThrGlyLeuValThrPheThrArgGln	700
Qy	2267	AGTCTCGAAGATTTCAGAAATGGAAAGATGTCAAAAAACCTTGACACCATTTGCATGTC	2326
Db	701	SerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHisVal	720
Qy	2327	ACTTACGAAGTACCATAGAGAAAAATGGCCACAGCATGTCTACAGGTGGATTTTGCAAT	2386
Db	721	ThrTyrGluGlyThrIleGluGluAsnGlyGlnGlyMetLeuGlnValAspPheAlaAsn	740
Qy	2387	CGTTTTGTGAGGTGGTGTAAACAGTCAGGACTTGTGCMAGAGAAATCCGCTTTTAA	2446
Db	741	ArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPheLeu	760
Qy	2447	ATCAATCTCGAGTTGATTATTTCACGGCTTTCTCACTGAGGTGCTGGATCACAATGAATG	2506
Db	761	IleAsnProGluLeuLeuIleIleSerArgLeuPheThrGluValLeuAspHisAsnGluCys	780
Qy	2507	CTAATTATCACAGGTACTGACAGGTACAGTGAATACACAGGCTATGCTGACATATCGT	2566
Db	781	LeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyrArg	800
Qy	2567	TGGTCCCGGAGCCAGAAAGATGGGAGTGAAGAGGACGACTCGCAGCGCGCTGCACCTGAG	2626

167 ATGAATCGGGCCCCGGCTGTGAAACCTCGACCAAAAGCGACCGCTGGGGCGCGCTACA 226  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db 1 MetSerAlaGlyProGlyCysGluProCysThrLysArgProArgTrpAspAlaAla 20  
  
227 ACTTCG---CCGGCTCTTCGGACGCCGGAGTTCCTCGAGCAGGCGAGGGCGCTC 283  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 21 ThrSerProProAlaAlaSerAspAlaArgSerPheProGlyArgGlnArgValLeu 40  
Db  
  
284 GACCCCAAGACCTCACGTGCAGTTCAGGGTCCACCGCTCTCGCGCAGCGCTCGCTCCCA 343  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 41 AspSerLysAspAlaProValGlnPheArgValProProSerSerSerGlyCysAlaLeu 60  
Db  
  
344 GGGCAGCGGGCAGCAGCAGAGCGCGCCACCTCGCTGTTTTCAAACAAAGACTATT 403  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 61 GlyArgAlaGlyGlnHisArgGlySerAlaThrSerLeuValPheLysGlnLysThrIle 80  
Db  
  
404 ACCAGTTCGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTGGATAGTAA 463  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 81 ThrSerTrpMetAspThrLysGlyIleLysThrValGluSerGluSerLysSerLys 100  
Db  
  
464 GAAAAACAACATACAGAATAGAAATCCATCATGATGAGTTCGTACAAAAGATAACTTTTAC 523  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 101 GluAsnAsnThrArgGluGluSerMetMetSerSerValGlnLysAspAsnPheTyr 120  
Db  
  
524 CAACATAATGTAGAAAATTAAGTAATGTTTCTCAGCTAAGTCTTGATTAAGTCACACT 583  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 121 GlnHisAsnMetGluLysLeuGluAsnValSerGlnLeuGlyPheAspLysSerProVal 140  
Db  
  
584 GAAAAAGTACACAGATTTGAAACAGCATCAGCTGCAGCAATGTTAAGTGGCAAAAT 643  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 141 GluLysGlyThrGlnLysLeuLysGlnHisGlnThrAlaAlaMetCysLysTrpGlnAsn 160  
Db  
  
644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGATAACCTGTGAPCCA 703  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 161 GluGlyProHisSerGluArgLeuLeuGluSerGluProProAlaValThrLeuValPro 180  
Db  
  
704 GAGCAGTTTAGTAATCTAACATTGATCGTCCGTACCTCAAAATGATGATCACAGTGACACA 763  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 181 GluGlnPheSerAsnAlaAsnValAspGlnSerSerProLysAspAspHisSerAspThr 200  
Db  
  
764 GATAGTGAAGAGAAATAGACAAATCAACAGTTTCTCACAACCTGTTAAAGCTGCAAAATGCA 823  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 201 AsnSerGluGluSerArgAspAsnGlnGlnPheLeuThrHisValLysLeuAlaAsnAla 220  
Db  
  
824 AAGCAGACTACGGAAGATCAACACCGCAGAGAGCCAAAGCCACCAAGAGTGCAGCAAG 883  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 221 LysGlnThrMetGluAspGluGlnGlyArgGluAlaArgSerHisGlnLysCysGlyLys 240  
Db  
  
884 TCTTGCCATCCTGGGGAAGACTGTGCAAGTTGTACAGCAAGATGAGATAGACGTGGTCCA 943  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 241 AlaCysHisProAlaGluAlaCysAlaGlyCysGlnGlnGluGluThrAspValValSer 260  
Db  
  
944 AAGAGTCCATGTTCAGATGTGGCTCTGAGGATGTTGGTACTGGGTCAAAAATGACAAC 1003  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 261 GluSerProLeuSerAspThrGlySerGluAspValGlyThrGlyLeuLysAsnAlaAsn 280  
Db  
  
1004 AAATTGATTAGCAAGAAAGTTCCTAGGAATTCCTCCATTTGAGAAAGAGAAAGTGAA 1063  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 281 ArgLeuAsnArgGlnGluSerSerLeuLysAsnSerProPheGluLysGluSerGlu 300  
Db  
  
1064 CCCGATACCGATGATGGATTAATCTTAAATAGTTGTCAAGACTCAGAGCAGAT 1123  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 301 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 320  
Db  
  
1124 GAGGAGACAAGTCCAGGTTTTCATGAACAAGAGATGGTAGTTCCTCCCAAAACAGCAAAAT 1183  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 321 GluGluThrSerProGlyPheAspGluGlnGluAspSerSerSerAlaGlnThrAlaAsn 340  
Db  
  
1184 AACCTTCAAGGTTCCAGCAGAGACGCTGACATTAATTAAGAAACGGTACTCTACT 1243  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 341 LysProSerArgPheGlnProArgGluAlaAspThrGluLeuArgLysArgSerAla 360  
Db  
  
1244 AAGGGCGGTGAAGTTAGATTACATTTCCATTTGAAGGAGGAGAGAGTGCCTGGAATG 1303  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY

361 LysGlyGluIleArgLeuHisPheGlnPheGluGlyGlySerArgAlaGlyMet 380  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
Db  
  
1304 AATGATTAAAGCTAAACCTACCTCGAATATTTCTAGCCTGAATGTAGAATGCAGAAAT 1363  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 381 AsnAspValAsnAlaLysArgProGlySerThrSerSerLeuAsnValGluCysArgAsn 400  
Db  
  
1364 TCTAAGCAACATCGGAAAAGGATCTTAAATACAGATCATTTGATGAGATGCCCAAA 1423  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 401 SerLysGlnHisGlyArgLysAspSerLysIleThrAspHisPheMetArgValProLys 420  
Db  
  
1424 CAGAGCAGCAGAAAGAAACAGTCGGGAAACAAACATCAACAGACACAGAAAGAGATC 1483  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 421 AlaGluAspLysArgLysGluGlnCysGluMetLysHisGlnArgThrGluArgLysIle 440  
Db  
  
1484 CTTAAATAGTTCACCTCACCTTCTCCAGATAAGAGTGGCTTGGAACTCCCATTTAG 1543  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 441 ProLysTyrIleProProHisLeuSerProAspLysLysTrpLeuGlyThrProIleGlu 460  
Db  
  
1544 GAGATGAGAAGAAATGCTCGGTGGATCCGGTGCCTCTCTTGAGACCATCTGCCAAT 1603  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 461 GluMetArgArgMetProArgCysGlyIleArgLeuProProLeuArgProSerAlaAsn 480  
Db  
  
1604 CACACAGTAAGTATTCGGGTAGATCTTTTGCAGCAGGAGAGTTCCTTAAACCTTTCCA 1663  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 481 HisThrValThrIleArgValAspLeuLeuArgIleGlyGluValProLysProPhePro 500  
Db  
  
1664 ACACATATAAGATTTGTGGATTAACAGCATGTTTAAATGCTTGTTCAGAACAAAAT 1723  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 501 ThrHisPheLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 520  
Db  
  
1724 TTGTACCCAGTCCGAGATGAGAAATGGTGAGCGAATCTGCGGGAGCCGGTGGAGCTCAT 1783  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 521 LeuTyrProValGluAspGluAsnGlyGluArgAlaAlaGlySerArgTrpGluLeuIle 540  
Db  
  
1784 CAGACTGCATCTCTCAACAAATTTACAGCACCCCAAACTTGAAGATGCTATTTCTGAAA 1843  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 541 GlnThrAlaLeuLeuAsnArgLeuThrArgProGlnAsnLeuLysAspAlaIleLeuLys 560  
Db  
  
1844 TACAATGTGGCATATTCTAAGAAATGGGACTTTTACAGCTTTTGTATTCGATTTCTGGGATAAG 1903  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 561 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuIleAspPheTrpAspLys 580  
Db  
  
1904 GTACTTGAAGAGCAGAGCTCAACATTTATATCATGTCCTCATCTGCTCATATGTGAAA 1963  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 581 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 600  
Db  
  
1964 AYTGCACCTGTCTGCCAAATTTTGCACCCAGCCAATACCACTCCTGAAACAGAGATG 2023  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 601 IleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGlnLysMet 620  
Db  
  
2024 AATCATTCATCAATGTTCGAGGAAACAGATTCGAGTTCCTTTAGCTAATCTTTCTTC 2083  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 621 AsnHisSerIleThrMetSerGlnGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 640  
Db  
  
2084 TCACATTCACAGCAGAAATCTAAGATGAATCGGAGTATTTAGTTACCCAGACATT 2143  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 641 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 660  
Db  
  
2144 AACTTCAATTCGATTTTGGAGGAGTTCATCAAGAAACCGGAGAACTTAAACCGCTC 2203  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 661 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 680  
Db  
  
2204 TTCTGCTACTTTAGAGATCACAGAAAAAACCTACTGGGTTGGTGACATTTACAAGA 2263  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 681 PheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPheThrArg 700  
Db  
  
2264 CAGACTCTTGAGATTTTCCAGAAATGGGAAGATCTGAAAACCTTGCACAGATTCAT 2323  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY 701 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysLeuLeuThrArgLeuHis 720  
Db  
  
2324 GTCACTTACAGAGTACCATAGAAGAAATGCCAAGGCATCTACAGGTGATTTTGA 2383  
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :  
QY

Db 721 ValThrTyrGluGlyThrIleGluGlyAsnGlyGlnGlyMetLeuGlnValAspPheAla 740  
 QY 2384 AATCGTTTGTGGAGTGGTGTACACAGTGCAGGACTTGTGCAGAGAAATCGCTTT 2443  
 Db 741 AsnArgPheValGlyGlyValThrSerAlaGlyLeuValGlnGluIleArgPhe 760  
 QY 2444 TTAATCAATCTGAGTGTATTTACAGGCTCTTCACTGAGTGTCTGTGATCACAATGAA 2503  
 Db 761 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 780  
 QY 2504 TGTCTTAATATCACAGGTACTGAGCAGTACAGTGAATACAGAGCTATGCTGAGACATAT 2563  
 Db 781 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 800  
 QY 2564 CGTTGGTCCGAGCCACGAAGATGGAGTGAAGGGACGACTGCGAGCGGCTGCACT 2623  
 Db 801 ArgTTPAlaArgSerHisGluAspArgSerGluArgAspTTPGlnArgThrThr 820  
 QY 2624 GAGATCGTTGCCATCGATGCTCTTCACTTCAGACCTACCTGATCAGTGTGTGCTGAG 2683  
 Db 821 GluIleValAlaIleAspAlaLeuHisPheArgTyrTyrLeuAspGlnPheValProGlu 840  
 QY 2684 AAAATGAGACGCGAGCTGAACAGGCTTACTGTGATTTCTCGTCTGGAGTTTCTTCA 2743  
 Db 841 LysIleArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValSerSer 860  
 QY 2744 GAGATCTTTCTGCGAGTGGCCACAGAACTGGGCTGTGGTCCCTTTGGGGGTGATGCC 2803  
 Db 861 GluAsnLeuSerAlaValAlaThrGlyAsnTTPGlyCysGlyAlaPheGlyAspAla 880  
 QY 2804 AGTTTAAAGCTTAAATACAGTATTTGCGAGTGTGCTGAGTGTGAGTGTGTTAT 2863  
 Db 881 ArgLeuLysAlaLeuIleGlnIleLeuAlaAlaValAlaGluArgAspValValTyr 900  
 QY 2864 TTCACCTTTGGGAGTCTCAGAATTGTATGAGACATTTTACAGCATGCACATTTTCTTACT 2923  
 Db 901 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 920  
 QY 2924 GAAAGAACTCACTGTGTGGAGATGTATAGCTGTGTGCTGATCATCATCATGAGAA 2983  
 Db 921 GluArgLysLeuThrValGlyValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 940  
 QY 2984 TGCAGAACTGTTCACCCCTGACACGACATCAAGCTTTATCCATTCATATACCATGCT 3043  
 Db 941 CysArgAsnGlySerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 960  
 QY 3044 CTCGAGTCTGTGTCAGACACCCCTGACCATTCAGGGCCAAAGGACAGGG 3091  
 Db 961 ValGluSerCysThrGlnThrThrAsnGlnProGlyGlnArgThrGly 976

## RESULT 5

O88622 PRELIMINARY; PRT; 968 AA.  
 AC O88622;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Poly(ADP-ribose) glycohydrolase.  
 GN PARG.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99380098; PubMed=10449915;  
 RA Ame J.C., Apion F., Jacobson E.L., Jacobson M.K.;  
 RT "Assignment of the poly(ADP-ribose) glycohydrolase gene (PARG) to  
 RT human chromosome 10q11.23 and mouse chromosome 14B by in situ  
 RT hybridization."  
 RL Cytogenet. Cell Genet. 85:269-270 (1999).  
 DR EMBL; AF079557; AAC28735.1; --  
 MG; MGI:1347094; Parg.

DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR InterPro; IPR007724; PARG.  
 DR Pfam; PF05028; PARG; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 968 AA; 109169 MW; 3981DA4809E1E007 CRC64;  
 Alignment Scores:  
 Pred. No.: 0 Length: 968  
 Score: 4395.00 Matches: 837  
 Percent Similarity: 90.99% Conservative: 52  
 Best Local Similarity: 85.67% Mismatches: 78  
 Query Match: 60.97% Indels: 10  
 DB: 11 Gaps: 6

US-09-302-812-3 (1-4069) x O88622 (1-968)

QY 167 ATGAATCGGGCCCCCGGCTGTGAACCTTCAGCAAGCGACCCGCTGGGGCGCGCT--- 223  
 Db 1 MetSerAlaGlyProGlyTTPGluProCysThrLysAla---ArgTTPGlyAlaAlaGly 19  
 QY 224 ACAACTTCGCGGCTGCTTCGGAGCGCCGAGCTTTCGAGCAGCAGCAGCGGCTCTC 283  
 Db 20 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 39  
 QY 284 GACCCCAAGACGCTCAGTGCAGTTCAGGTCGCCACCGTCTCGCAGCCTGGCTCCCA 343  
 Db 40 AspProLysAspAlaProValGlnPheArgValProProSerProAlaCysValSer 59  
 QY 344 GGGCAGCGCGGACACAGCAGAGGCGGCCACCTCGCTGTTGTTTCAACAAAGACTATT 403  
 Db 60 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 79  
 QY 404 ACCAGTTGATGACACTTAAGGAATCAAGACCGGAATCAGAAAGTTTGGATAGTAAA 463  
 Db 80 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 96  
 QY 464 GAAACCAACATACAGAATAGATCATGATGATGTTCTGTACAAAAGATAACTTTTAC 523  
 Db 97 GluAsnAsnThrArgLysAspSerMetSerSerValGlnLysAspAspPheTyr 116  
 QY 524 CAACATATGTAGAAAAATTAGTAAATGTTTCTCAGCTAAGCTTTGATGATCACTACT 583  
 Db 117 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 136  
 QY 584 GAAAAAGTACACAGTATTTGAACACGATCAGCTGCGCAATCTGTAGTGCAGAAAT 643  
 Db 137 GluLysSerSerGlnTyrLeuAsnGlnGlnGlnThrAlaSerValCysLysTyrProGlnAsn 156  
 QY 644 GAAGGGAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTGTGATCCA 703  
 Db 157 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 176  
 QY 704 GAGCAGTTTACTAATGCTAACATTCATGCTGGTCACTCAAAATGATGATCAGTGCACA 763  
 Db 177 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 196  
 QY 764 CATAGTGAAGAAATAGAGACAATCAACAGTTTCTCAACTGTAAAGCTTGCAAAATGCA 823  
 Db 197 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 216  
 QY 824 AAGCAGACTACGGAAGATGAACACGCGCAGAGAGCAAAAGCCACAGAGTGCAGCAAG 883  
 Db 217 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 233  
 QY 884 TCTTGCCATCTCGGGAGACTGTGCAAGTTGTGCAAGATGATGATGATGATGATGATG 943  
 Db 234 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluGluValAspValLeuPro 253  
 QY 944 AAGAGTCCATTGTGATGTTGGCTCTGAGGATGTTGGTACTGGTCAAAAATGACAAAC 1003  
 Db 254 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 273  
 QY 1004 AAATTGATTAGACAAAGAAAGTTGCTAGGAAATCTCTCCATTTGAGAGAGAAAGTGA 1063

Db 274 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 293  
 QY 1064 CCGGAATCACCAGTGGATGTTCTAAATAATAGTTGTCAAGACTCAGAAGCAGAT 1123  
 Db 294 ProGluSerProMetAspValAspAsnSerArgAsnSerCysGlnAspSerGluAlaAsp 313  
 QY 1124 GAGAGACAAGTCCAGGTTTGTAGTGAACAAGAGATGGTAGTCTCTCCCAAGCAAAAT 1183  
 Db 314 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 332  
 QY 1184 AAACCTTCAGGTTCCAGCAAGAGACGCTGACATTCGAATTTAGGAACGGTACTCTACT 1243  
 Db 333 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 352  
 QY 1244 AAGGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGAGAGATCGCACTGGAATG 1303  
 Db 353 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 371  
 QY 1304 AATGATTAAATGCTAAACTACCTCGGAATATTTCTAGCCTGTAATGTAGAAATGCAAAAT 1363  
 Db 372 SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 391  
 QY 1364 TCTAAGCAACATCGAAGAAAGGATCTTAAATACAGATCATTTGTAGACTGCCCAAA 1423  
 Db 392 SerLysGlnHisGlyLysArgSerLysIleThrAspHisPheMetArgIleSerLys 411  
 QY 1424 GCAGAGACAGAGAAAAGAACAGTGGGAAACCAACATCAAGAAACAGAAAGGAAGATC 1483  
 Db 412 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 431  
 QY 1484 CCTAAATACGTTCCACTCACTTCTCCAGATAAGAGTGGCTTGGAACTCCCATGAG 1543  
 Db 432 ProLysTyrIleProProAsnLeuProGluLysLysTrpLeuGlyThrProLysGlu 451  
 QY 1544 GAGATGAGAAGATCGCTCGGTGGATCGCGTGGCTCTCTTGACCATCTGCCAAT 1603  
 Db 452 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 471  
 QY 1604 CACACAGTAACTATTCGGGTAGATCTTTTCGAGCAGGAGAGTTCCTAAACCTTTTCCA 1663  
 Db 472 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 491  
 QY 1664 ACATTTAAAGATTTGTGGATTAACAAGCATGTTAAATAGCTTGTTCAGAACAAAT 1723  
 Db 492 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 511  
 QY 1724 TTCTACCATGAGATGAGATGGTGGAGCACTGCGGGAGCGGTGGAGCTCAT 1783  
 Db 512 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 531  
 QY 1784 CAGACTGCATCTCTCAACAAATTTACACGACCCCAAACTTGAAGGATGCTATTCTGAAA 1843  
 Db 532 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 551  
 QY 1844 TACAATGTGCATTTCTAAGAAATGGGACTTTACAGCTTTGTATCGATTCTCGGATAAG 1903  
 Db 552 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 571  
 QY 1904 GTACTTTGAAGCAAGCAAGCTCAACATTTATATCAGTCCATCTTGCTGATATGGTGAAA 1963  
 Db 572 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 591  
 QY 1964 ATTGCATCTGTCTGCCAAATTTTGCACCCAGCAATACCACTCTCTGAAACAGAAAGATG 2023  
 Db 592 IleAlaLeuCysLeuProAsnIleCysThrGlnProLysLeuLeuLysGlnLysMet 611  
 QY 2024 AATCATTCATCAATGTGCGAGGAACAGATTCGCAGTCTTTTAGCTAATGCTTTCTTC 2083  
 Db 612 AsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAlaPhePhe 631  
 QY 2084 TGCACATTTCCAGCAAGATGCTAGATCAATCGAGTATTTCTAGTTACCCAGACATT 2143

Db 632 CysThrPheProArgArgAsnAlaLysMetLysSerGluTyrSerSerTyrProAspIle 651  
 QY 2144 AACTTCAATCGATTGTTGAGGACGCTTCATCAAGGAAACCGGAGAACTTAAACCGTC 2203  
 Db 652 AsnPheAsnArgLeuPheGluGlyArgSerSerArgLysProGluLysLeuLysThrLeu 671  
 QY 2204 TTCTGCTACTTTAGAGAGTCAAGAGAAACCACTACTGGTTGGTGCATTTACAGA 2263  
 Db 672 PheCysTyrPheArgArgValThrGluLysProThrGlyLeuValThrPheThrArg 691  
 QY 2264 CAGAGCTTGAAGATTTCCAGAAATGGGAAGATGTGAAAAACCTTGACACCATTCAT 2323  
 Db 692 GlnSerLeuGluAspPheProGluTrpGluArgCysGluLysProLeuThrArgLeuHis 711  
 QY 2324 GTCACCTTACGAAGCTACCATAGAGAAATGGCCAAAGCATGCTACAGGTGGATTTCGA 2383  
 Db 712 ValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAspPheAla 731  
 QY 2384 AATCGTTTTTGGAGGTGGTGAACACAGTGCAGGACTTGTGCAAGAAAGAAATCCGCTTT 2443  
 Db 732 AsnArgPheValGlyGlyValThrGlyAlaGlyLeuValGlnGluGluIleArgPhe 751  
 QY 2444 TTAATCAATCCTCAGTGTGATTTTTCACGCTCTTCACTGAGTGTGATCACAATGAA 2503  
 Db 752 LeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHisAsnGlu 771  
 QY 2504 TGTCTAATATATCAGCTACTGAGCAGTAGTACAGTGAATACACAGGCTATGCTGAGACATAT 2563  
 Db 772 CysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGluThrTyr 791  
 QY 2564 CGTGTGTCGAGCAGCAGCAAGATGGAGTGAAGGAGCACTGCGAGCGCGCTGCACT 2623  
 Db 792 ArgTrpAlaArgSerHisGluAspGlySerGluLysAspTrpGlnArgArgCysThr 811  
 QY 2624 GAGATCGTTCGCACTGCTCTTCACTTCACTGAGCTTCACTGAGTGTGCTGCTGAG 2683  
 Db 812 GluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheValProGlu 831  
 QY 2684 AAATGAGAGCGCAGCTGAACAGGCTTACTGTGGATTTTCCGTCCTCGAGTGTCTTCA 2743  
 Db 832 LysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyValProSer 851  
 QY 2744 GAGAACTTCTGCACTGCGCAGCAAGAACTGGGCTGTGCTGCTTGGGCTGATGCC 2803  
 Db 852 GluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGlyAspAla 871  
 QY 2804 AGGTAAAAAGCCTTAATACAGATATTGGCAGCTGCTGAGCTGAGAGATGTGGTTAT 2863  
 Db 872 ArgLeuLysAlaLeuIleGlnIleLeuAlaIleAlaAlaGluArgAspValValTyr 891  
 QY 2864 TTCACCTTTGGGACTCAGAAATTGATGAGACATTTACAGCATGCAATTTTCTTACT 2923  
 Db 892 PheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPheLeuThr 911  
 QY 2924 GAAAGAAAACCTCACTCTGAGATGTATAGCTGTCTGATGATGATGATGATGATGATG 2983  
 Db 912 GluArgLysLeuAspValGlyLysValTyrLysLeuLeuLeuArgTyrTyrAsnGluGlu 931  
 QY 2984 TGCAAGAACTGTTTCCACCTCGCAGCAGCATCAAGCTTTATCCATTTCATATACCATGCT 3043  
 Db 932 CysArgAsnCysSerThrProGlyProAspIleLysLeuTyrProPheIleTyrHisAla 951  
 QY 3044 GTCAGTCTCTGTCAGAGACCGCTGACCATTCAGGCAAGGAGAGGAGGACC 3094  
 Db 952 ValGluSerSerAlaGluThrThrAspMetProGlyGlnLysAlaGlyThr 968

RESULT 6

Q80YQ6 PRELIMINARY; PRT; 961 AA.  
 ID Q80YQ6; AC Q80YQ6; DT 01-JUN-2003 (TEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)

DE poly (ADP-ribose) glycohydrolase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC050892; AAH50892.1; -.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
DR InterPro; IPR007724; PARG.  
DR Pfam; PF05028; PARG; 1.  
KW Hydrolase.  
SQ SEQUENCE. 961 AA; 108580 MW; 1558F120C367CF63 CRC64;  
Alignment Scores:  
Pred. No.: 0 Length: 961  
Score: 4381.50 Matches: 833  
Percent Similarity: 91.12% Conservative: 50  
Best Local Similarity: 85.96% Mismatches: 77  
Query Match: 60.79% Indels: 9  
DB: 11 Gaps: 5  
US-09-302-812-3 (1-4069) x Q80YQ6 (1-961)  
QY 167 ATGAATGGGGCCCGCTGTGAACCTTCACCAAGCGCCCGCTGGGGCGCGCT--- 223  
DB 1 MetSerAlaGlyProGlyTrpGluProCysThrLysArgProArgTrpGlyAlaAlaGly 20  
QY 224 ACAACTTCGCGCGCTTCGAGACCGCGGACTTCCGAGCAGGCGCGCTCCTC 283  
DB 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 40  
QY 284 GACCCCAAGGCGCTCAGTCAGTTCAGGTTCACCGTCCACCGTCTCCGCGCGCTCCCA 343  
DB 41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60  
QY 344 GGGCAGGCGGACACACAGCAGCGCCACTCGCTGTTTTCACCAAGACTATT 403  
DB 61 GlyArgAlaGlyProHisArgGlyAsnAlaThrSerPheValPheLysGlnLysThrIle 80  
QY 404 ACCAGTTGATGGACACTAAAGGAATCAAGACAGCGGAATCAGAAAGTTTGATAGTAAA 463  
DB 81 ThrThrTrpMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97  
QY 464 GAAACAAACATACAGATAGATCCATGATGAGTTCGTGTCACAAAGATACCTTTTAC 523  
DB 98 GluAsnAsnAsnThrArgIleAspSerMetMetSerValGlnLysAspAsnPheTyr 117  
QY 524 CAACATAATGTAGAAAAATTAGTAATGTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583  
DB 118 ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137  
QY 584 GAAAAAGTACACAGTATTGAAACGACATCAGACTGCAGCAATGTGTAAGTGGCAAAAT 643  
DB 138 GluLysSerSerGlnTyrLeuAsnGlnGlnInThrAlaSerValCysLysTrpGlnAsn 157  
QY 644 GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAACAGTAAACCTCGGTACCA 703  
DB 158 GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177  
QY 704 GAGCATTTAGTAATGCTAACATTTGATCGTCCACTCAAAATGATGATCAGAGTCACACA 763  
DB 178 LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197  
QY 764 GATAGTGAAGAAATAGACACATCAACAGTATTCTCACAACCTGTAAAGCTTCGAATGCA 823  
DB 198 AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProIleLysLeuAlaAsnThr 217  
QY 824 AAGCAGACTACGGAAGATGAACACGCGCAGAGGCCAAAGCCACAGAGTGCAGCAG 883

DB 218 LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 234  
QY 884 TCTTGCCATCTCTGGGGAAGACTGTGCAAGTTGTGAGCAAGATGAGATAGAGCTGGTCCA 943  
DB 235 SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAspValLeuPro 254  
QY 944 AAGAGTCCATTGTGAGTGTGGCTGTGAGATGTTGCTACTGGTCAAAATAATGACAAAC 1003  
DB 255 GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274  
QY 1004 AAATTGATTAGCAAGAAAGTTCCTCCTCAGCAATTTCTCCTCCTCCTCCTCCTCCTCCT 1063  
DB 275 LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProPheGluLysGluSerGlu 294  
QY 1064 CCCGAATCACCGATGATGTGATAATTTCTAAAAATAGTTGTCAAGACTCAGAGACGAT 1123  
DB 295 ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314  
QY 1124 GAGGACAGATCCAGGTTTGTGATGAACAAGAGATGTAGTTCTCTCCCAACAGCAAT 1183  
DB 315 GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 333  
QY 1184 AAACCTTCAGGTTCCAGCAAGAGACGCTGACATTTGAATTTAGCAACGCTACTCTACT 1243  
DB 334 LysLeuSerSerCysGlnAlaArgGluAlaAspGlyLeuArgLysArgTyrLeuThr 353  
QY 1244 AAGGCGGTGAAGTTAGATTACATTTCCAAATTTCAAGGAGGAGAGAGTCGCACTGGAATG 1303  
DB 354 LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnValGlyThr 372  
QY 1304 AATGATTTAAATGCTAAACTACCTGGAAATTTCTAGCCCTGAATGTAGATGCAAGAAAT 1363  
DB 373 SerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValGluCysArgSer 392  
QY 1364 TCTAAGCAACATGAAAGAAAGATTTCTAAATTCACAGATCATTTGATGAGACTGCCCAA 1423  
DB 393 SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysLys 412  
QY 1424 GCAGAGACAGAGAAAGAAAGACAGTGGGAAACCAACAAACATCAAGAACAGAAAGATC 1483  
DB 413 SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432  
QY 1484 CTTAAATAGCTTCCACCTCCTTCCAGATAGAGTGGCTTGGAACTCCCACTTCCAG 1543  
DB 433 ProLysTyrIleProProAsnLeuProProGluLysLysTrpLeuGlyThrProIleGlu 452  
QY 1544 GAGATGAGAAGATGCTCGGTGGGATCCGGTCCCTCTCTTGAGACCATCTGCCAAT 1603  
DB 453 GluMetArgLysMetProArgCysGlyIleHisLeuProSerLeuArgProSerAlaSer 472  
QY 1604 CACACAGTAACTATTTCGGGTAGATCTTTTGGCAGCAGAGAGTTCCTAAACCTTTTCCA 1663  
DB 473 HisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysProPhePro 492  
QY 1664 ACACATTAATAGATTTGGGGATAACAAGCATGTTAAATGCCCTTGTTCAGAACAAAT 1723  
DB 493 ThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGluGlnAsn 512  
QY 1724 TTGTACCCAGTGGGAAGTGAAGATGGTGGAGCAACTCGGGGAGCGGTGGGAGCTCAAT 1783  
DB 513 LeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGluLeuIle 532  
QY 1784 CAGACTGCACCTCTCAACAAATTTACAGACCCCAAACTTGAAGGATGCTATTCTGAAA 1843  
DB 533 GlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIleLeuLys 552  
QY 1844 TACATGTGGCATATTCTTAAGAAATGGCACTTTACAGCTTTGATGCTTCTGGGATAG 1903  
DB 553 TyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrpAspLys 572  
QY 1904 GTACTTGAAGAGCAGAGAGCTCAACATTTATATCAGTCCATCTTCCCTGTGATGTGAAA 1963  
DB 573 ValLeuGluGluAlaGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMetValLys 592



QY	644	GAAGGAAACACACGAGCAGCTTTTGGAAAGTGAACCTCAAAACAGTAACTGGTACCA	703	QY	1718	CAAAATTTGTACCCAGTGAAGATGAGAAATGGTGAAGCACTGGGGAGCGGTGGAG	1777
Db	158	AlaGlyGluArgAlaGluGlnLeuSerAlaSerGluProSerAlaValThrGluAlaPro	177	Db	514	GlnAsnLeuTyrProValGluAspGluAsnGlyGluArgThrAlaGlySerArgTrpGlu	533
QY	704	GAGCAGTTTAGTAATGCTAAACATTCATCGGTCACTCAAAATGATGATGATCAGTGACACA	763	QY	1778	CTCATTTCAGACTGCTACTTCTCAACAAATTTACACACCCCAAAACTTGAAGGATCTATT	1837
Db	178	LysGlnLeuSerAsnAlaAsnIleAspGlnSerProProThrAspGlyHisSerAspThr	197	Db	534	LeuIleGlnThrAlaLeuLeuAsnLysPheThrArgProGlnAsnLeuLysAspAlaIle	553
QY	764	GATAGTGAAGAAATAGAGACAAATCAACAGTTTCTCAACCTGTAAGCTGGCAATGCA	823	QY	1838	CTGAAATACAAATGTGGCATATTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTGG	1897
Db	198	AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProValLysLeuAlaAsnAla	217	Db	554	LeuLysTyrAsnValAlaTyrSerLysLysTrpAspPheThrAlaLeuValAspPheTrp	573
QY	824	AAGCAGACTACGGAAGATGAACACACCCAGAGAAAGCCACCAAGTGCAGCAAG	883	QY	1898	GATTAAGGTACTTGAAGAGCAGAGCTCAACATTATATATCATGCTCATCTTGCCTGATATG	1957
Db	218	LysGlnThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerAla	234	Db	574	AspLysValLeuGluGluAlaGlnHisLeuTyrGlnSerIleLeuProAspMet	593
QY	884	TCTTGCCATCTCTGGGGAAGACTGTGCAAGTTCTCGCAAGATGATGATGATGCTGCA	943	QY	1958	GTGAAATTTGCACCTGCTCTGCCAAATATTTTGCCACCCAGGCCAAATACCACTCTCTGAAACAG	2017
Db	235	SerCysGlnCysGlyGlnAspCysAlaGlyCysGlnArgGluGluAlaAspValIlePro	254	Db	594	VallysIleAlaLeuCysLeuProAsnIleCysThrGlnProIleProLeuLeuLysGln	613
QY	944	AGAGTCCATTTCTCAGATGTGGCTCTGAGGATGTGGTACTGGCTCAAAAATGACAC	1003	QY	2018	AAGATGAATCATTCATCCATCCATCTCGCAGGAACAGATTGCCAGTCTTTTACGTAATGCT	2077
Db	255	GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlySerLysAsnAspAsn	274	Db	614	LysMetAsnHisSerValThrMetSerGlnGluGlnIleAlaSerLeuLeuAlaAsnAla	633
QY	1004	AAATTTGATTAGACAAAGATTGCTTAGGAAATTTCTCTCCATTTGAGAAGGAAAGTGAA	1063	QY	2078	TTCTTCTGCACATTTCCACGCAAAATGCTTAAGATGAATCGGAGTATTCTAGTTACCCA	2137
Db	275	LysLeuThrGlyGlnGluSerGlyLeuGlyAspSerProPheGluLysGluSerGlu	294	Db	634	PhePheCysThrPheProArgAsnAlaLysMetLysSerGluTyrSerSerTyrPro	653
QY	1064	CCGAAATCACCAGTGGATGTGAATATCTTAAATATAGTTGTCAAGACTCAGAACAGAT	1123	QY	2138	GACATTAATCTCAATCGATTGTTGAGGGAGCTTCATCAAGAAACCCGAGAAACTTAAA	2197
Db	295	ProGluSerProMetAspValAspAsnSerLysThrSerCysGlnAspSerGluAlaAsp	314	Db	654	AspIleAsnPheAsnArgLeuPheGluArgSerSerArgLysProGluLysLeuLys	673
QY	1124	GAGGACACAGCTCCAGTTTTCATGACAAA-----GAAGATGGTAGTTCTCCCAACA	1177	QY	2198	ACGCTCTTCTGCTACTTTAGAAAGAGTCAACAGAGAAAAACCTACTGGGTGGTGACATT	2257
Db	315	GluGluAlaSerProValPheAspGluGlnAspAspGlnAspAspArgSerSerGlnThr	334	Db	674	ThrLeuPheCysTyrPheArgArgValThrGluLysLysProThrGlyLeuValThrPhe	693
QY	1178	GCAATTAACCTTCAAGGTTCAAGCAAGAGACGCTGACATTAATTTAGGAAACGGTAC	1237	QY	2258	ACAAGACAGAGTCTTGAAGATTTTCCAGAAATGGGAAAGATGTGAAACCCCTTGACACA	2317
Db	335	AlaAsnLysLeuSerSerArgGlnAlaArgGluValAspGlyAspLeuArgLysArgTyr	354	Db	694	ThrArgGlnSerLeuGluAspPheProGluTrpGluArgCysAspLysProLeuThrArg	713
QY	1238	TCTACTAAGGGCGGTGAAGTTAGATTACATTTCCAAATTTGAAGGAGGAGAGTGGCACT	1297	QY	2318	TTGCATGTCTACTTCAAGAGGTACCATAGAATAATGCCAAGGCATGTCACAGGTGGAT	2377
Db	355	LeuThrLysGlySerGluIleArgLeuHisPheGlnPheGluGlyGly---SerAsnAla	373	Db	714	LeuHisValThrTyrGluGlyThrIleGluGlyAsnGlyArgGlyMetLeuGlnValAsp	733
QY	1298	GGAATCAATGATTAAATCTTAATCTTAACTGCTGAAATATTTCTAGCTGAATGTAGAATGC	1357	QY	2378	TTTGCAAAATCGTTTTTGGAGGTGGTGTAAACAGTCCAGGACTTGTGCAGAGAAATC	2437
Db	374	GlyThrSerAspLeuAsnAlaLysProSerGlyAsnSerSerLeuAsnValAspGly	393	Db	734	PheAlaAsnArgPheValGlyGlyGlyValThrGlyAlaGlyLeuValGlnGluIle	753
QY	1358	AGAAATTTCTAAGCAACATCGAAAAAGGATTTCTAAATACACAGATCATTTGATGAGACTG	1417	QY	2438	CGCTTTTAAATCAATCCTGAGTTGATTATTTACCGCTCTTCCACTGAGTGTGATCAC	2497
Db	394	ArgSerSerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheValArgIle	413	Db	754	ArgPheLeuIleAsnProGluLeuIleValSerArgLeuPheThrGluValLeuAspHis	773
QY	1418	CCCAAGCAGAGCAGAGAAAGAAACAGTGGGAAACCAACATCAAGAACAGCAAGAGG	1477	QY	2498	AATGAATGTCTAATTTATCACAGGTACTGACAGTACAGTGAATACACAGGTATGCTGAG	2557
Db	414	ProLysSerGluAspLysArgLysGluGlnCysGluValArgHisGlnArgAlaGluArg	433	Db	774	AsnGluCysLeuIleIleThrGlyThrGluGlnTyrSerGluTyrThrGlyTyrAlaGlu	793
QY	1478	AGATCCCTTAATACGTTCCACTCACCTTTCTCCAGATAAAGAGTGGCTTGGAACTCCC	1537	QY	2558	ACATATCGTTGGTCCGAGCCACCAAGATGGAGTGAAGAGGACGACTGCAGCGGCGC	2617
Db	434	LysIleProLysTyrValProProAsnLeuProProAspLysLysTrpLeuGlyThrPro	453	Db	794	ThrTyrArgTrpAlaArgSerHisGluAspGlySerGluLysAspAspTrpGlnArgCys	813
QY	1538	ATTGAGGAGATGAGAAGATGCTCGGTGGGATCCGGCTGCTCTCTTGAGACCATCT	1597	QY	2618	TGCATCTGAGATCGTTGCCATCGATCTCTTCACTTTCAGACGCTACTCATCAGTTTGTG	2677
Db	454	IleGluGluMetArgLysMetProArgCysGlyValArgLeuProLeuLeuArgProSer	473	Db	814	CysThrGluIleValAlaIleAspAlaLeuHisPheArgArgTyrLeuAspGlnPheVal	833
QY	1598	GCAATCACACAGTAACTATTCGGGTAGATCTTTTCGGAGCAGGACAGTCTCTAACT	1657	QY	2678	CCTGAGAAATGAGACGCGAGCTGAACCAAGGCTTACTGTGGATTTCTCGTCTCGAGTT	2737
Db	474	AlaSerHisThrValThrValArgValAspLeuLeuArgAlaGlyGluValProLysPro	493	Db	834	ProGluLysValArgArgGluLeuAsnLysAlaTyrCysGlyPheLeuArgProGlyVal	853
QY	1658	TTTCCACACATTAATAAGATTGTGGGATAACAACATGTTTAAATGCTTGTTCAGAA	1717	QY	2738	TCTTCAGAGATCTTTTTCAGTGGCCACAGAAACTGGGGCTGTGGTCTTGGGGT	2797
Db	494	PheProThrHisTyrLysAspLeuTrpAspAsnLysHisValLysMetProCysSerGlu	513	Db	854	ProProGluAsnLeuSerAlaValAlaThrGlyAsnTrpGlyCysGlyAlaPheGlyGly	873
				QY	2798	GATGCCAGTTTAAAGCCCTTAATACAGATATTGGCAGCTGTGCTGAGCGAGATGNG	2857



```

|||||
874 AspalargLeuLysAlaLeuLleGlnLeuLeuAlaAlaAlaAlaGluArgAspVal 893
|||||
2858 GTTATTATTCACCTTGGGACTCAGAAATGATGAGAGACATTTACAGCATGCGACATTTTC 2917
|||||
894 ValTyrPheThrPheGlyAspSerGluLeuMetArgAspIleTyrSerMetHisThrPhe 913
|||||
2918 CTTACTCAAGAGGAACCTCAGTGTGGAGATGCTGTATAGCTGTGTACCATCTACAAT 2977
|||||
914 LeuThrGluArgLysLeuAsnValGlyLysValTyrArgLeuLeuLeuArgTyrTyrArg 933
|||||
2978 GAAGAAATGACAGAACTGTTCCACCCCTGGACACACATCAAGCTTTTATCCATTCAATATAC 3037
|||||
934 GluGlnCysArgAspCysSerSerProGlyProAspThrLysLeuTyrProPheIleTyr 953
|||||
3038 CATGCTCTCAGAGTCTCTGTCAGAGACCGCTGACCATTCAGGGCAAGAGCAGGGACC 3094
|||||
954 HisAlaAlaGluSerSerAlaGluThrSerAspGlnProGlyGlnArgThrGlyThr 972
|||||

RESULT 8
Q8CB72 PRELIMINARY; PRT; 920 AA.
ID AC Q8CB72;
DC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Poly.
GN PARG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR ENBL; AK036656; BAC29519.1; -.
DR MGD; MGI:1347094; Parg.
DR InterPro: IPR007724; PARG.
DR Pfam; PF05028; PARG; 1.
SQ SEQUENCE 920 AA; 103799 MW; 379E779E121B09A1 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 920
Score: 4165.50 Matches: 793
Percent Similarity: 90.95% Conservative: 51
Best Local Similarity: 85.45% Mismatches: 75
Query Match: 57.79% Indels: 9
DB: 11 Gaps: 5

US-09-302-812-3 (1-4069) x Q8CB72 (1-920)
QY 167 ATGAATCGGGCCCCCGCTGTGAACCTCGACCAAGGACCCCGCTCGGGGCGCGCT--- 223
Db 1 MetSerAlaGlyProGlyTyrGluProCysThrLysArgProArgTyrGlyAlaAlaGly 20
QY 224 ACACTTCGGCGCTGCTTCGGACCGCGGAGCTTCGGACAGCGAGCGCGCTCCTC 283
Db 21 ThrSerAlaProThrAlaSerAspSerArgSerPheProGlyArgGlnArgValLeu 40
QY 284 GACCCCAAGGACGCTCAGTTCAGGTTCACCGTCCACCGTCTCCGCGACCGCTCGCTCCA 343
Db 41 AspProLysAspAlaProValGlnPheArgValProProSerSerProAlaCysValSer 60
QY 344 GGGCAGCGGGACGACGACAGGAGCGGCGACCTCGCTGTTGTTTCAACAAAGACTATT 403
Db 61 GlyArgAlaGlyProHisArgLysAsnAlaThrSerPheValPheLysGlnLysThrIle 80

```

```

QY 404 ACCAGTTGGATGACACTAAAGGAATCAAGACAGCGGAATCAGAAAAGCTTGGATAGTAAA 463
Db ||||| ThrThrTyrMetAspThrLysGlyProLysThrAlaGluSerGlu-----SerLys 97
QY 464 GAAAAACAATACAGAAATAGAAATCCATGATGAGTCTGTACAAAAAGATAACTTTTAC 523
Db ||||| GluAsnAsnThrArgIleAspSerMetSerValGlnLysAspAsnPheTyr 117
QY 524 CAACATAATCTAGAAAAATAGTAAATGTTCTCAGCTAAGTCTTGATAAGTCACTCACT 583
Db ||||| ProHisLysValGluLysLeuGluAsnValProGlnLeuAsnLeuAspLysSerProThr 137
QY 584 GAAAAAGTACACAGATTTTGAACCAAGCATCAGACATGACCAATGTGTAAAGTGGCAAAAT 643
Db ||||| GluLysSerSerGlnTyrLeuAsnGlnGlnThrAlaSerValCysLysTyrGlnAsn 157
QY 644 GAGGGGAAAACACCGGAGCAGCTTTTGGAAAGTGAACCTCAAAACAGTAACCTCGGTACCA 703
Db ||||| GluGlyLysHisAlaGluGlnLeuLeuAlaSerGluProProAlaGlyThrProLeuPro 177
QY 704 GAGCGTTTAGTAATGCTAACATTGATCGCTCACCTCAAAATGATGATCAAGTGACACA 763
Db ||||| LysGlnLeuSerAsnAlaAsnIleGlyGlnSerProHisThrAspAspHisSerAspThr 197
QY 764 GATAGTCAAGAGAAATAGAGCAATCAACAGTCTTCTCAAACTGTAAAGCTTGCAAAATGCA 823
Db ||||| AspHisGluGluAspArgAspAsnGlnGlnPheLeuThrProLysLysLeuAlaAsnThr 217
QY 824 AAGCAGACTACGGAAGATGAACACGCGCAGAGAGCCAAAGCCACCAAGAGTGCAGCAAG 883
Db ||||| LysProThrValGlyAspGlyGlnAlaArg-----SerAsnCysLysCysSerGly 234
QY 884 TCTTGCCATCTCGGAGACTGTGCAAGTGTGCAAGATGTGTCAGCAAGATGATGATGAGTGTGCGCA 943
Db ||||| SerArgGlnSerValLysAspCysThrGlyCysGlnGlnGluValAlaAspValLeuPro 254
QY 944 AAGAGTCCATTGTGAGATTTGGCTCTGAGGATGTTGCTACTGCTGCTCAAAAATGACAAC 1003
Db ||||| GluSerProLeuSerAspValGlyAlaGluAspIleGlyThrGlyProLysAsnAspAsn 274
QY 1004 AAATTGATTAGACAAGAAAGTTCCTCAGGAAATTCCTCTCATTTGAGAGGAAAGTGAA 1063
Db ||||| LysLeuThrGlyGlnGluSerSerLeuGlyAspSerProProPheGluLysGluSerGlu 294
QY 1064 CCGCAATCACCGATGGATGTGATTAATCTCAAAATAGTGTCAAGACTCAGAGCAGAT 1123
Db ||||| ProGluSerProMetAspValAspAsnSerLysAsnSerCysGlnAspSerGluAlaAsp 314
QY 1124 GAGGACAAAGTCCAGGTTTGTATGAACAAGAGATGTTCTCTCCCAACAGCAAAAT 1183
Db ||||| GluGluThrSerProValPheAspGluGlnAspAsp---ArgSerSerGlnThrAlaAsn 333
QY 1184 AAACCTTCAAGGTTCCAAGCAAGAGCGCTGACATTCGAATTTAGGAAACGGTACTCTACT 1243
Db ||||| LysLeuSerSerCysGlnAlaArgGluAlaAspGlyAspLeuArgLysArgTyrLeuThr 353
QY 1244 AAGGGCGTCAAGTACATTTCAATTTGAAGGAGAGAGACTCGCAGCTCGGAATG 1303
Db ||||| LysGlySerGluValArgLeuHisPheGlnPheGlu---GlyGluAsnAsnAlaGlyThr 372
QY 1304 AATGATTAAATGCTAAACTACCTCGGAATATTTCCTAGCCTGAATGTAGAAATGCAAGAAAT 1363
Db ||||| SerAspLeuAsnAlaLysProSerGlyAsnSerSerSerLeuAsnValGluCysArgSer 392
QY 1364 TCTAAGCAACATCGAAAAAGGATTTCAAAATCACAGATCATTTGTAGAGACTGCCCAAA 1423
Db ||||| SerLysGlnHisGlyLysArgAspSerLysIleThrAspHisPheMetArgLysSerLys 412
QY 1424 GCAGGACGACAGAAAGAACAGAGTGGGAAACCAACATCAAAAGAACAGAAAGAGATC 1483
Db ||||| SerGluAspArgArgLysGluGlnCysGluValArgHisGlnArgThrGluArgLysIle 432
QY 1484 CCTAAATACGTTCCACCTCACCTTTCTCCAGATAAGAAAGTGGCTTGGAATCCCATTTGAG 1543

```











Db	408	AlaSerHisArgThrProLeuAsnAspValGluMetAsnArgGluLeuProAlaAsnAsn	427	Db	117	---GlnLysAlaAspHisValLeuAspGlyValLysSerGlyLeuArgLeuLeuGlyPro	135
Qy	2729	-----CCTGGAGTTCTTCAGAGATCTT-----TCTGCAGTG	2761	Qy	2030	-----TCCATCAACAATCGCAGAAACAGATTGCGAGTCTTTTATGCT	2077
Db	428	LeuLeuArgAspPheTyrValGluGlyValAspAsnGluAspHisGluAspGlyVal	447	Db	136	GlnGluAlaGlyIleValLeuLeuSerGlnGlnLeuIleAlaLeuLeuAlaCysSer	155
Qy	2762	CCACAGGAACCTGGGCTGGTGGCTGGTGGGCTGATGCCAGGTAAAGCCCTTAATA	2821	Qy	2078	TTCTTCGCACATTTCCACGACGAAATGCTAAGATGAATCGGAGTATCTTCTGTACCCA	2137
Db	448	AlaThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysAlaThrIle	467	Db	156	PhePheCysLeuPheProGluValAspArgSerLeuLys-----AsnLeuGln	171
Qy	2822	CAGATATTGGCAGCTGCTGACGCTGAGCGAGATGGTT---TATTACCTTTGGGAC	2878	Qy	2138	GACATTAACTTCAATCGATTCTTTGAG-----GGACGTTTCATCAAGGAAACCG	2185
Db	468	GlnTrpLeuAlaAlaSerGlnThrArgArgProPheIleSerTyrThrPheGly---	486	Db	172	GlyIleAsnPheSerGlyLeuPheSerPheProTyrMetArgHisCysThrLysGlnGlu	191
Qy	2879	TCAGATTCATGACAGACATTTACAGCATGACATTTCTTACTGAAAGAACTCAT	2938	Qy	2186	GAGAACTTAAACCGCTCTCTGCTACTTTAGAGAGTCACAGAGAAAACCTACTCGG	2245
Db	487	ValGluAlaLeuArgAsnLeuAspGlnValThrLysTrpIleLeuSerHisLysTrpThr	506	Db	192	AsnLysIleLysCysLeuIleHisTyrPheGlyArgIleCysArgTirPheProThrGly	211
Qy	2939	GTTGGAGATGTATTAAGCTGTGTGCTACGATAC	2971	Qy	2246	TGGTGACATTTACAGACACAGAGCTTGAA-----GATTTT	2281
Db	507	ValGlyAspLeuTrpAsnMetMetLeuGluTyr	517	Db	212	PheValSerPheGluArgLysIleLeuProLeuGluTyrHisProHisPheValSerTyr	231
RESULT 12				Qy	2282	CCAGAA-----TGGGAAAGATGTCAAAAACCCCTTGACACAGTTCATGCTACTTAC	2332
Q8VY1		PRELIMINARY; PRT; 522 AA.		Db	232	ProLysAlaAspSerTrpAlaAsnSerValThrProLeuCysSerIleGluIleHisThr	251
AC	Q8VY1			Qy	2333	GAAAGTACCATAGAAATAATGGCCAGGCATGCTACAGGTGGATTTTGCATTCGTTT	2392
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			Db	252	SerGlyAlaIleGluAspGlnProCysGluAlaLeuGluValAspPheAlaAspGluTyr	271
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			Qy	2393	GTTGAGGTGGTGAACAGTCAGGACTTGTGCAAGAGAAATCCGCTTTTATCAAT	2452
DE	Putative poly(ADP-ribose) glycohydrolase.			Db	272	PheGlyGlyLeuThrLeuSerTyrAspThrLeuGlnGluGluIleArgPheValIleAsn	291
GN	AT2G31870.			Qy	2453	CCTGAGTTGATTATTTCACGGCTCTTCTACCTGAGGTGCTGATCAATGAATGTCTAAT	2512
OS	Arabidopsis thaliana (Mouse-ear cress).			Db	292	ProGluLeuIleAlaGlyMetIlePheLeuProArgMetAspAlaAsnGluAlaIleGlu	311
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			Qy	2513	ATCACAGGTACTGAGCAGTACAGTGAATACACAGGCTACTGTCAGACATATCGTTGGTCC	2572
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			Db	312	IleValGlyValGluArgPheSerGlyTyrThrGlyTyrGlyProSerPheGlnTyrAla	331
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			Qy	2573	CGGAGCCAGAGATGGGATGCAAGGGACGACTGCGAGCGCTGCTGCTGAGATCGTT	2632
OX	NCBI_TaxID=3702;			Db	332	GlyAspTyrThrAspAsnLysAspLeuAspIlePheArgArgGlyThrArgValIle	351
RN	[1]			Qy	2633	GCCATCGATGCTCTTTCATCTTACAGCTACCTGATGATGTTTGTGCTGAGAAATGAGA	2692
RA	SEQUENCE FROM N.A.			Db	352	AlaIleAspAlaMetPro---AspProGlyMetGlyGlnTyrLysLeuAspAlaLeuIle	370
RA	Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,			Qy	2693	CGGAGCTGAACAGGCTTACTGTGATTTCTCCGT-----	2728
RA	Palm C.J., Bowser L., Jones T., Banu J., Carninci P., Chen H.,			Db	371	ArgGluValAsnLysAlaPheSerGlyTyrMetHisGlnCysLysTyrAsnIleAspVal	390
RA	Chen R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,			Qy	2729	-----CCTGGAGTTTCTTCA-----	2743
RA	Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,			Db	391	LysHisAspProGluAlaSerSerHisValProLeuThrSerAspSerAlaSerGln	410
RA	Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,			Qy	2744	-----GAGAACTCTTCTGAGTGGCC	2764
RA	Ecker J., Theologis A., Davis R.W.			Db	411	ValIleGluSerSerHisArgTrpCysIleAspHisGluGluLysLysIleGlyValAla	430
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			Qy	2765	ACAGGAACTGGGGCTGGTGGCTTTGGGGTGGATGCCAGGTAAAGCTTAAATACAG	2824
DR	EMBL; AY072330; AAL61937.1; -			Db	431	ThrGlyAsnTrpGlyCysGlyValPheGlyGlyAspProGluLeuLysIleMetLeuGln	450
DR	GO; GO:0016787; F:hydrolase activity; IEA.			Qy	2825	ATATTGGCAGCTCTGCTGAGCTGAGCGAGATGCTGGT---TATTTCACCTTTGGGGACTCA	2881
DR	InterPro; IPR007724; PARG.			Db	451	TrpLeuAlaIleSerGlnSerGlyArgProPheMetSerTyrTyrThrPheGly---Leu	469
DR	Pfam; PF05028; PARG; 1.			Qy	2882	GAATTGATGAGACATTTTACAGCATGCACATTTCTTCTTACTGAAAGAACTCCTGTT	2941
KW	Hydrolase.			Db	470	GlnAlaLeuGlnAsnLeuAsnGlnValIleGluMetValAlaLeuGlnMetThrVal	489
SQ	SEQUENCE 522 AA; 58942 MW; 6FF98685D867CA15 CRC64;						
Alignment Scores:							
Pred. No.:	4,62e-34	Length:	522				
Score:	539.00	Matches:	133				
Percent Similarity:	49.65%	Conservative:	82				
Best Local Similarity:	30.72%	Mismatches:	138				
Query Match:	7.48%	Indels:	80				
DB:	10	Gaps:	11				
US-09-302-812-3 (1-4069) x Q8VY1 (1-522)							
Qy	1892	TTCTGGGATAGGTACTTGAAGAGCAGACGCTCAACATTTATATCATGCTCATCTTGCCT	1951				
Db	80	PhePheAspLysLysIleSerArgGluGluSerAlaAsnPhePheGlyGluValValPro	99				
Qy	1952	GATATGGTGAATATGCACTCTGCTGCCAATATTTGGCACCACCCAGCCCACTCCTG	2011				
Db	100	AlaLeuCysArgLeuLeuLeuGlnLeuProSerMetLeuGluLysHisTyr---	116				
Qy	2012	AAACAGAGATGAATCAT-----	2029				



Db	367	euhisaen-MetTyrGluPheAspProAspIleThrTyrIleLeuProAlaLeuGluMet	386
QY	1934	TATCAGTCCATCTTCCTGATATGCTG	1960
Db	387	TyrTyrIysGluMetSerGluLeuValGlyArgGluValLeuGluLysPheAlaArg	406
QY	1961	-----AAAATTTGCACTCTGCTGCCAAATATTTGCCACCCAGCAATACCACTCTCTGAAA	2014
Db	407	ValAlaArgIleAlaLysThrAlaGluAspIleLeuProGluArgIleTyrArgLeuVal	426
QY	2015	CAGAAGATGAATCAATTCATCAATGTCGAGCAAGATGTCGAGTCTTTTACGTAAT	2074
Db	427	GlyAspVal---GluSerAlaThrLeuSerHisLysGlnCysAlaAlaValAlaArg	445
QY	2075	GCTTTCTTCGACATTTCCACGACGAAATGCTAAGATGAATCGAGTATTTCTAGTTAC	2134
Db	446	MetPhePhe-----AlaArgProAspSerProPheSer-----	456
QY	2135	CCAGACATTAATCTCAATTCGATGTTTGGAGGAGCTTCATCAAGGAACCGAGAACTT	2194
Db	457	-----PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu	472
QY	2195	AAAAGCTCTTCTGCTACTCTTAGAGACTCAGAGAAAAAACCCTACTGCGTGGTGACA	2254
Db	473	LysPheLeuPheThrTyrPheAspLysMetSerMetAspProProAspGlyAlaValSer	492
QY	2255	TTT-----ACAAGACAGAGTCTTGAAGATTTT---CCAGAAATGGGAAAGATGTGAAAA	2305
Db	493	PheArgLeuThrLysMetAspLysAspThrPheAsnGluGluTyrLys-----AspLys	510
QY	2306	CCCTTGACAGATTCGATGCTACT-----TAGAAGGTACATAGAGAAATGCGCAA	2359
Db	511	LysLeuArgSerLeuProGluValGluPhePheAspGluMetLeuIleGluAspThrAla	530
QY	2360	GGCATGCTACAGTGGATTTTCAATCTGTTGTTGAGGTGGTGTAAACGATCGACGA	2419
Db	531	LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGly	550
QY	2420	CTTTGCAAGAGAATCCGCTTTTAAATCTCGAGTTGATTAATTTACGGCTCTTC	2479
Db	551	SerValGlnGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeuLeu	570
QY	2480	ACTGAGTGTGATCAATCAATGCTAAATATCATCAGGTACTGAGCAGTACAGTGAA	2539
Db	571	CysGluLysMetIleGlnLeuAlaIleSerIleValGlyAlaTyrValPheSerSer	590
QY	2540	TACACAGGCTATGCTGAGACATATCTGCTG-----TCCCGGAGCCACGAA	2584
Db	591	TyrThrGlyTyrGlyHisThrLeuLysTyrAlaGluLeuGlnProAsnHisSerArgGln	610
QY	2585	GATGGAGTGAA---AGGACGACGTCGAGCGCGCTGCACTGAGATCCTTCCATCGAT	2641
Db	611	AsnThrAsnGluPheArgAspArgPheGlyArgLeuArgValGluThrIleAlaIleAsp	630
QY	2642	GCCTCTCACTTCAGACG-----TACCTGCATCAGTTTGTGCTGAGAAA	2686
Db	631	AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn	650
QY	2687	ATGAGCGGAGCTGAACAGGCTTACTGTGGATTTCTCGCTCCTCGAGTTTCTTCAGAG	2746
Db	651	IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr	670
QY	2747	AATCTTTCTCAGTGGCCACAGGAAATCGGGCTGTGGTGCCTTTGGGGGTGATGCCAGG	2806
Db	671	AsnIle---ProIleValThrGlyTyrTipGlyCysGlyAlaPheAsnGlyAspLysPro	689
QY	2807	TTAAAGCCTTAATACAGATTTGGGAGCTGCTGCAGCTCAGCGAGATGTGGTTATTTTC	2866
Db	690	LeuLysPheIleIleGlnValIleAlaAlaGlyValAlaAspArgProLeuHisPheCys	709
QY	2867	ACCTTTGGGACTCAGAAATTTGATGAGAGACATTTTACAGCATGCACATTTTCTTACTGAA	2926
Db	710	SerPheGlyGluProGluLeuAlaAlaLysCysLysLysIleIleGluArgMetLysGln	729
QY	2927	AGAAACTCAGCTCTGAGATGTATAGCTGTCTGTCTAGCATACTACAATGAAGATGC	2986
Db	730	LysAspValThrLeuGluGlyMetLeuPheSerMetIle-----	741
QY	2987	AGAAACTGTTTCCACCCCTGGACGACATCAACGCTTTATCCATTC	3031
Db	742	AsnAsnThrGlyLeuProHisLysHisPheGluPheTyrValPhe	756
RESULT 14			
ID	Q19637	PRELIMINARY;	PRT; 764 AA.
AC	Q19637;	01-NOV-1996 (TrEMBLrel. 01, Created)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	PME-3 protein (corresponding sequence F20C5.1) (Poly ADP-ribose metabolism enzyme-3 short form).		
GN	F20C5.1 OR PME-3.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	SEQUENCE FROM N.A.		
RA	Mathews P.		
RL	Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=99069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C.elegans: A platform for investigating biology."		
RL	Science 282:2012-2018(1998).		
RN	SEQUENCE FROM N.A.		
RA	Gagnon S.N., Hardy I., Desnoyers S.;		
RT	"Characterization of poly(ADP-ribose) glycohydrolases in the nematode Caenorhabditis elegans."		
RL	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; Z68161; CAA92299.2; -.		
DR	EMBL; AY185494; AAC26317.1; -.		
DR	PIR; T21138; T21138.		
DR	WormPep; F20C5.1b; CR32867.		
DR	InterPro; IPR007724; PARG.		
DR	Pfam; PF05028; PARG; 1		
SQ	SEQUENCE 764 AA; 87441 MW; B391703B1D0CB79A CRC64;		
Alignment Scores:			
Pred. No.:	9,776-29	Length:	764
Score:	473.00	Matches:	211
Percent Similarity:	37.88%	Conservative:	128
Best Local Similarity:	23.58%	Mismatches:	290
Query Match:	6.56%	Indels:	266
DB:	5	Gaps:	38
US-09-302-812-3 (1-4069) x Q19637 (1-764)			
QY	599	TATTTGAACCACTCAGACTCGACCAATGTGTAAGTGGCAAAATGAAGGAAACACACG	658
Db	27	PheAlaHisGluValProThrMetLysArgLysLeuThrGluHisGlyAsnThrThr	46
QY	659	GAGCAGCTTTTGGAAAGTGAACCTCAACAGTAACCTGGTACCGAGCAGCTTTTAGTAAT	718
Db	47	GluSerLysGluAsp-----ProGluGluProLysSer	57
QY	719	GCTAACATGATCGGTTCACCTCAAAATGATGATCAGTCAGACATAGTAGAGAAAT	778
Db	58	ArgAspValPheValSerSerGlnSerSerAspGluSerGlnGluAspSerAlaGluAsn	77
QY	779	-----AGAGACAATCAACAGTTTCTCACA---ACTGTAAAGCTT	814



Db 78 ProGluIleAlaLysGluValSerGluAsnCysGluAsnLeuThrGluThrLeuLysIle 97  
QY 815 GCAAAATGCAAGCAG-----ACTACGGAAGATGACACACCCAGAGAGCCAAA 862  
Db 98 SerAsnIleGlnSerLeuLeuAspAsnValThrGluArgSerGluHisThrLeuAsp----- 115  
QY 863 AGCCACACAGAAAGTGCAGCAAGTCTTCGCATCTCTGGGAGAGACTGTGCAAGTTCACGAA 922  
Db 116 AsnHis-----LysSerThrGluProMetGluGlu----- 125  
QY 923 GATGAGATGACGTGGTGCACCAAGAGATCCATTTGTCAGATCTTGGCTCTGAGGATGTTGGT 982  
Db 126 -----AspValAsnAsnLysSerAsnIle-----AspValAla 136  
QY 983 ACTGGGTCAAAAATGACACAAATTCATGATGACAGAAAGTTCGCCTAGGAAATCTCCT 1042  
Db 137 IleAsnSerAspGluAspGluLeuValLeuGluGluAsn----- 150  
QY 1043 CCATTTGAGAAGGAAGTGAACCCGAATCACCGATGAGTGGATTAATCTAAATAATAGT 1102  
Db 151 -----AsnLysGluMetArg-----AspGlyGluGlnValGlnLeu 163  
QY 1103 TGTCAAGACTCAGAGCAGATGAGGAG-----ACAAGTCCAGGTTTGTGATGACAA 1153  
Db 164 SerGlnAspLeuPheAlaAspAspGlnGluLeuIleGluTyrProGlyIleMetLysAsp 183  
QY 1154 GAA-----GATGTAGTTCCTCCCAACACAGCAATAAACCTTCAAGG 1195  
Db 184 ThrThrThrGlnLeuAspIleThrAspSerGluValGluThrAlaGlnLysMetGluMet 203  
QY 1196 TTCCAAGCAAGACGCTGCATTTGAATTT-----AGGAACCG 1234  
Db 204 IleGluGluThrGluAlaAspSerThrPheValGlyGluAspSerLysAlaThrLysThr 223  
QY 1235 TACTCTACTAAGGCG----- 1249  
Db 224 ValArgThrSerSerSerPheLeuSerThrValSerThrCysGluAlaProAlaLys 243  
QY 1250 GGTGAAGTTAGATTACAT-----TTCCAATTTGAAGGAGGA 1285  
Db 244 GlyArgAlaArgMetTyrGlnLysGluLeuGluLysHisValIleAlaPheThrGluGly 263  
QY 1286 GAGAGTCGCACTGCAATGAATGATTTAAATGCTAAACTACTCGA----- 1330  
Db 264 ---AsnLeuThrLeuGlnProAspLeuAsnLysValAspProAspArgAsnTyrArgTyr 282  
QY 1331 -----ANATTTCTAGCTGAATGATGATGAGTGAATGCAAAATCTAAGCACATGGA 1378  
Db 283 CysThrIleProAsnPheProAlaSerGlnGlyLysLeuArgGluAspAsnArgTyrGly 302  
QY 1379 AAAAAGGATTTAAATACAGATCAATTTGATGAGCTGCCAAAGACAGAGACAGAGA 1438  
Db 303 ProLys-----IleValLeuPro-----GlnArg 310  
QY 1439 AAGAACAGCTGGGAACCAACATCAAGAACAGAA-----AGGAAG 1480  
Db 311 TrpArgGluPheAspSerArgGlyArgArgAspSerTyrPheTyrPheLysArgLys 330  
QY 1481 ATCCCTAAATACGTTCCACCTCACCCTTCTCCAGATGAAGTGGCTTGGACTCCCAT 1540  
Db 331 LeuAspGlyTyrLeuLysCysTyrLysThrThrGlyTyrPheMetPheValGlyLeuLeu 350  
QY 1541 GAGGAGATGAGAAGATG---CCTCGGTGTGGATCCGGCTGCTCTCTTGGAGACCATCT 1597  
Db 351 HisAsnMetTrpGluPheAspProAspIleThrTyrLysLeuProAlaLeuGlu----- 368  
QY 1598 GCCAATCACACAGTAATTTCCGGGTAGATCTTTTTCGACGAGGAGAGTTCCTAAACCT 1657  
Db 368 ----- 368  
QY 1658 TTTCACACACATTAAGATTTCTGGGATGAACAAGCATGTTAAATGCTTGTTCAGAA 1717  
Db 369 -----MetTyrTyrLysGluMet----- 374

QY 1718 CAAAATTTGTACCCAGTGAAGATGAGAAATGTTGACGGAACCTGCGGGAGCCGGTGGAG 1777  
Db 375 -----SerGlu 376  
QY 1778 CTCATT---CAGACTGCATCTTCTCAACAAATTTTACACGACCCCAAACTTGAAGGATGCT 1834  
Db 377 LeuValGlyArgGluGluValLeuGluLysPheAlaArgValAlaArgIleAla----- 394  
QY 1835 ATTCTGAAATACAATGTGGCATATTTCTAAGAAATGGGACTTTACAGCTTTGATCGATTTC 1894  
Db 394 ----- 394  
QY 1895 TGGGATPAAGTACTTGAAGAAGCAGAAAGTCAACATTATATATATATATATATATATATATAT 1954  
Db 395 -----LysThrAlaGluAspIleLeuProGluArgIleTyrArg---LeuValGlyAsp 411  
QY 1955 ATGGTCAAAATTCGACTCTGTCTGCCAAATATTTGACCCAGCCCAATACCACTCTCGAAA 2014  
Db 412 Val----- 412  
QY 2015 CAGAAGATGAATCATTCATCCATCACAATGTGCAGAAACAGATTGCCAGTCTTTTAGCTAAT 2074  
Db 413 -----GluSerAlaThrLeuSerHisLysGlnCysAlaAlaLeuValAlaArg 428  
QY 2075 GCTTTCTTCTGCAATTTCCACGACGAAATGTGAAGTGAATCGAGATTTCTAGTTAC 2134  
Db 429 MetPhePhe-----AlaArgProAspSerProPheSer----- 439  
QY 2135 CCAGACATTAATTCATTCATCGATTGTTGGGACGTTTCATCAGGAAACCGGAGAACTT 2194  
Db 440 -----PheCysArgIleLeuSerSerAspLysSerIleCysValGluLysLeu 455  
QY 2195 AAAACGCTCTTCGTCTACTTTAGACAGTGCACAGAGAAAAACCTACTGGTGTGGTGA 2254  
Db 456 LysPheLeuPheThrTyrPheAspLysMetSerMetAspProAspGlyAlaValSer 475  
QY 2255 TTT-----ACAAGACAGAGTCTTTGAAGATTTT---CCAGAATGGGAAAGATGTGAAA 2305  
Db 476 PheArgLeuThrLysMetAspLysAspThrPheAsnGluLysLys-----AspLys 493  
QY 2306 CCCTTGACAGTGTGATGTCATCT-----TACGAAGTTACCATAGAGAAATGGCCAA 2359  
Db 494 LysLeuArgSerLeuProGluValGluPheAspGluMetLeuLeuGluAspThrAla 513  
QY 2360 GGCATGCTACAGTGCATTTGCAATCGTTTGTGGAGTGTGTAAACAGTGCAGGA 2419  
Db 514 LeuCysThrGlnValAspPheAlaAsnGluHisLeuGlyGlyValLeuAsnHisGly 533  
QY 2420 CTTGTCAAGAGAAATCCGCTTTTAAATCAATCCTGAGTGTGATTTTCCCGCTCTTC 2479  
Db 534 SerValGlnGluIleArgPheLeuMetCysProGluMetMetValGlyMetLeu 553  
QY 2480 ACTGAGTGTGATCACAATGATGTCTAATATATACAGTACTGACAGTACAGTGA 2539  
Db 554 CysGluLysMetLysGlnLeuGluAlaIleSerIleValGlyAlaTyrValPheSerSer 573  
QY 2540 TACACAGGCTATCTGAGACATATCGTTGG-----TCCCGAGCCAGCA 2584  
Db 574 TyrThrGlyTyrGlyHisThrLeuLysTyrAlaGluLeuGlnProAsnHisSerArgGln 593  
QY 2585 GATGGAGTGA---AGGACGAGCTGCGAGCGCGCTGCAGCTGAGATCGTTCGATCGAT 2641  
Db 594 AsnThrAsnGluPheArgAspPheGlyArgLeuArgValGluThrIleAlaIleAsp 613  
QY 2642 GCTCTTCATCTCAGACGC-----TACCTGATCAGTTTGTGCTGAGAAA 2686  
Db 614 AlaIleLeuPheLysGlySerLysLeuAspCysGlnThrGluGlnLeuAsnLysAlaAsn 633  
QY 2687 ATGAGACGCGAGCTGAACAGGCTTACTGTGAGTTTCTCGTCTCTCGGATTTCTTCAGAG 2746  
Db 634 IleIleArgGluMetLysLysAlaSerIleGlyPheMetSerGlnGlyProLysPheThr 653





[illegible]

Search completed: May 26, 2004, 17:08:53  
Job time : 291.952 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 17:19:25 ; Search time 111.821 Seconds  
(without alignments)  
2466.140 Million cell updates/sec

Title: US-09-302-812-4  
Perfect score: 5190  
Sequence: 1 MNAGPCFCTKATWGAAT.....YHVESCAETADHSQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5190	100.0	976	5	Aae25630 Human pol
2	5190	100.0	976	5	Aau76021 Human pol
3	5190	100.0	976	5	Aau76012 Human pol
4	5190	100.0	976	6	Abg72279 Human pol
5	4635.5	89.3	977	5	Aae25629 Bovine po
6	4635.5	89.3	977	5	Aau76020 Bovine po
7	4635.5	89.3	977	5	Aau75799 Bovine po
8	4635.5	89.3	977	6	Abg72278 Bovine po
9	4395	84.7	968	5	Aae25631 Murine po
10	4395	84.7	968	5	Aau76022 Mouse pol
11	4395	84.7	968	5	Aau76013 Mouse pol
12	4395	84.7	968	6	Abg72280 Murine p
13	1078	20.8	768	4	Abbs9491 Drosophil
14	1078	20.8	768	5	Aae25632 Fruit fly
15	1078	20.8	768	5	Aau76023 Fruit fly
16	1078	20.8	768	5	Aau76014 Fruit fly
17	1078	20.8	768	6	Abg72281 Fruit fly
18	463	8.9	726	5	Aae25633 Poly aden
19	463	8.9	726	5	Aau76024 Worm poly
20	463	8.9	726	5	Aau76015 Worm poly
21	463	8.9	726	6	Abg72282 C. elegan
22	344.5	6.6	819	4	Abg20721 Novel hum
23	210	4.0	100	4	Abg11103 Novel hum
24	178	3.4	33	5	Aae25652 Bovine po
25	174	3.4	4873	6	Abol14747 Novel hum

26	173.5	3.3	748	4	AAM58340	Human bra
27	170	3.3	335	4	ABG09631	Novel hum
28	170	3.3	2703	4	ABB60074	Drosophil
29	167	3.2	31	5	Aae25634	Bovine po
30	167	3.2	31	5	Aae25651	Bovine po
31	167	3.2	31	5	Aau76025	Bovine po
32	167	3.2	31	5	Aau76016	Bovine po
33	167	3.2	31	6	ABG72283	Oligoscept
34	167	3.2	1274	6	ABU18980	Pathogen
35	167	3.2	1274	6	ABM71209	Staphyloc
36	167	3.2	1803	4	ABb65391	Drosophil
37	166.5	3.2	1163	4	Aau28028	Novel hum
38	166.5	3.2	2194	4	AAM40114	Human pol
39	165.5	3.2	1274	6	ABU15913	Protein e
40	164.5	3.2	1111	4	AAU37362	Staphyloc
41	162	3.1	768	4	AAU78686	Human pro
42	160	3.1	1212	4	ABG04004	Novel hum
43	160	3.1	1213	4	ABG23201	Novel hum
44	160	3.1	1213	4	ABG18490	Novel hum
45	160	3.1	1241	4	ABG14773	Novel hum

ALIGNMENTS

RESULT 1  
AAE25630  
ID AAE25630 standard; protein; 976 AA.

XX AAE25630;

XX 04-NOV-2002 (first entry)

XX Human poly adenosine diphosphate-ribose glycohydrolase (PARG).

XX Human; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
XX ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neotropic;  
XX cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
XX Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
XX cytoskeletal; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
XX antisense therapy.

OS Homo sapiens.

XX US6395543-B1.

XX 28-MAY-2002.

XX 23-FEB-2000; 2000US-00511507.

XX 01-MAY-1998; 98US-0083768P.

XX 30-APR-1999; 99US-00302812.

XX (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;

XX WPI; 2002-535641/57.

XX N-PSDB; AAD42082.

XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-ribose glycohydrolase involved in cellular response to DNA damage, inhibition of which is useful for treating neoplastic disorders and neurodegenerative diseases.

XX Claim 3; Col 55-60; 77pp; English.

XX The invention relates to an isolated nucleic acid molecule which encodes a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG) which catalyses release of ADP-ribose from an ADP-ribose polymer. The invention is useful as probes and primer molecules that can be used in hybridisation assays and polymerase chain reaction (PCR) amplification. The knowledge of the nucleotide sequence of the PARG gene permits the

CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury, Parkinson's  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is human PARG  
 XX  
 SQ Sequence 976 AA;

Query Match 100.0%; Score 5190; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNAGPCPCCTKATWGAATTPAASDARSPPRRVLDPKDAHVQFRVPPSPACVPG 60  
 DB 1 MNAGPCPCCTKATWGAATTPAASDARSPPRRVLDPKDAHVQFRVPPSPACVPG 60

QY 61 QAGQHRGSATSLVFKQKTTISMDTKGKTAESLDSKNNNTRIESMSSVQKDNFYQ 120  
 DB 61 QAGQHRGSATSLVFKQKTTISMDTKGKTAESLDSKNNNTRIESMSSVQKDNFYQ 120

QY 121 HNVKLVNVSQSLDKSLTEKSTQVILNQHOTAAMCKWQEGKHTQILLESPTQVTLVPE 180  
 DB 121 HNVKLVNVSQSLDKSLTEKSTQVILNQHOTAAMCKWQEGKHTQILLESPTQVTLVPE 180

QY 181 QFSNANIDRSPPNDHSDTDSEENRDNQOFTTVKLANAKOTTEDEHAREAKSHKCSKS 240  
 DB 181 QFSNANIDRSPPNDHSDTDSEENRDNQOFTTVKLANAKOTTEDEHAREAKSHKCSKS 240

QY 241 CHPGEDCASCQDEIDVVPKSLPSVGSSEIDVTGSKNDKILQBSCLGNSPPFFKESEP 300  
 DB 241 CHPGEDCASCQDEIDVVPKSLPSVGSSEIDVTGSKNDKILQBSCLGNSPPFFKESEP 300

QY 301 ESPMDVNSKNSQDSEADEETSPGDEQDSSQTANKPSRFQARDADIEFRKRYSTK 360  
 DB 301 ESPMDVNSKNSQDSEADEETSPGDEQDSSQTANKPSRFQARDADIEFRKRYSTK 360

QY 361 GGEVRLHFOFEGESBRTGMNDLNKLPNGNISLVNVECRNSKQHGKDKSKITDHLMLPKA 420  
 DB 361 GGEVRLHFOFEGESBRTGMNDLNKLPNGNISLVNVECRNSKQHGKDKSKITDHLMLPKA 420

QY 421 EDRRKEQWETKHQTERKTPKVVPKSLPSVGSSEIDVTGSKNDKILQBSCLGNSPPFFKESEP 480  
 DB 421 EDRRKEQWETKHQTERKTPKVVPKSLPSVGSSEIDVTGSKNDKILQBSCLGNSPPFFKESEP 480

QY 481 TVTIRVDLLRAGEVKKPPTTHYKQKLDWKNKVMPCSEQNLVPEDENGERTAGSRWELIQ 540  
 DB 481 TVTIRVDLLRAGEVKKPPTTHYKQKLDWKNKVMPCSEQNLVPEDENGERTAGSRWELIQ 540

QY 541 TALLNKFTROPNLKDAILKYNVAYSKKWDFTALIDFWKVLBEAEQAHLQSIPLDMVKI 600  
 DB 541 TALLNKFTROPNLKDAILKYNVAYSKKWDFTALIDFWKVLBEAEQAHLQSIPLDMVKI 600

QY 601 ALCLPNICTOPTPLKQKNHSTMSQEQIASILANAFCTPPRRNAKMSYSSYPDIN 660  
 DB 601 ALCLPNICTOPTPLKQKNHSTMSQEQIASILANAFCTPPRRNAKMSYSSYPDIN 660

QY 661 FNRLPEGRSSRPEKLTLCFYFRRVTEKPTGLVTFQSLDEDPWERCCKPLTRLHV 720  
 DB 661 FNRLPEGRSSRPEKLTLCFYFRRVTEKPTGLVTFQSLDEDPWERCCKPLTRLHV 720

QY 721 TYEGTTEENGQGLQVDFANRFVGGGVTAGLQVEEIRFLINPELLIISRLTEVLDPHNEC 780  
 DB 721 TYEGTTEENGQGLQVDFANRFVGGGVTAGLQVEEIRFLINPELLIISRLTEVLDPHNEC 780

QY 781 LIITGTEQSYETGYAETTRWSHSDGSDRDCERRCTTEIIVADLHFRYLDQVPEK 840  
 DB 781 LIITGTEQSYETGYAETTRWSHSDGSDRDCERRCTTEIIVADLHFRYLDQVPEK 840

QY 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 900  
 DB 841 MRRELKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 900

QY 901 TFGDSELMRDYISWHIFLTERKLTGVDVYKLLRLRYNEECNCSCTPGPDINKLYPFIYHAV 960  
 DB 901 TFGDSELMRDYISWHIFLTERKLTGVDVYKLLRLRYNEECNCSCTPGPDINKLYPFIYHAV 960

QY 961 ESCAETADHSGQRTGT 976  
 DB 961 ESCAETADHSGQRTGT 976

RESULT 2  
 AAU76021  
 ID AAU76021 standard; protein; 976 AA.  
 XX  
 AC AAU76021;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Human; poly(ADP-ribose) glycohydrolase; PARG; DNA repair; ADP-ribose;  
 KW adenosine diphosphate-ribose; neoplastic disorder; adenocarcinoma;  
 KW leukaemia; lymphoma; melanoma; myeloma; sarcoma; teratocarcinoma;  
 KW hyperplasia; hypertrophy; reperfusion; ischaemia; heart attack; stroke;  
 KW neurodegenerative disease; neurological disorder; Alzheimer's disease;  
 KW Huntington's disease; Parkinson's disease.  
 XX  
 OS Homo sapiens.  
 XX  
 FN US6337202-B1.  
 XX  
 PD 08-JAN-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511477.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI; 2002-163240/21.  
 DR N-PSDB; ABK14932.  
 XX  
 PT Novel isolated and purified poly (ADP-ribose) glycohydrolase protein  
 PT which catalyses release of ADP-ribose from ADP ribose polymer, useful for  
 PT treating neoplastic and neurological disorders, heart attack and stroke.  
 XX  
 PS Claim 2; Col 55-60; 81pp; English.  
 XX  
 CC The present invention relates to a new poly(ADP-ribose) glycohydrolase  
 CC (PARG) protein which catalyses release of ADP-ribose from an ADP  
 CC (adenosine diphosphate)-ribose polymer. The PARG molecule of the  
 CC invention is useful for generating antibodies and can be inhibited or  
 CC activated for diagnosing and treating neoplastic disorders such as  
 CC adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,  
 CC teratocarcinoma, hyperplasia and hypertrophy, reperfusion following  
 CC ischaemia, heart attack, stroke, neurodegenerative diseases, neurological  
 CC disorders including Alzheimer's, Huntington's and Parkinson's diseases,  
 CC and related conditions. PARG levels may be enhanced to suppress DNA  
 CC repair and increase the cell's susceptibility to chemotherapy drugs.  
 CC Antagonists of PARG are administered to treat or prevent neoplastic  
 CC disorders. The present amino acid sequence represents the human PARG  
 CC protein of the invention. This protein is one of several PARG proteins  
 CC (AAU76020-AAU76024) of the invention  
 XX  
 SQ Sequence 976 AA;

Query Match 100.0%; Score 5190; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNAGPGCEPCTKATRWGAATTSAAADARSFPQRORVLPDKAHVQFRVPPSPACVPG 60  
 DB 1 MNAGPGCEPCTKATRWGAATTSAAADARSFPQRORVLPDKAHVQFRVPPSPACVPG 60  
 QY 61 QAGHGRGATSLVFKQKTIISWMDTKGKTAESESLDSKENNTRIESMSSVQKDNFYQ 120  
 DB 61 QAGHGRGATSLVFKQKTIISWMDTKGKTAESESLDSKENNTRIESMSSVQKDNFYQ 120  
 QY 121 HNVKLVNVSQSLDKSLTEKSTQYLNQHOTAAAMCKWQNEGKTEQLLESEPQTIVLVE 180  
 DB 121 HNVKLVNVSQSLDKSLTEKSTQYLNQHOTAAAMCKWQNEGKTEQLLESEPQTIVLVE 180  
 QY 181 QFSNANIDRSPQNDHSDTSEENRDNQFLTTVKLANAKQTTEDEHAREAKSHOKCSKS 240  
 DB 181 QFSNANIDRSPQNDHSDTSEENRDNQFLTTVKLANAKQTTEDEHAREAKSHOKCSKS 240  
 QY 241 CHGEDCASCQDEIDVVPKSPISDVGSBDVGTSGKNDKLIROESCLGNSPPEKESEP 300  
 DB 241 CHGEDCASCQDEIDVVPKSPISDVGSBDVGTSGKNDKLIROESCLGNSPPEKESEP 300  
 QY 301 ESPMDVNSKNSCODSEADSETSPGDEQDGGSSQTANKPSRFOARDADIEFRKRYSTK 360  
 DB 301 ESPMDVNSKNSCODSEADSETSPGDEQDGGSSQTANKPSRFOARDADIEFRKRYSTK 360  
 QY 361 GGEVRLHFQEGGESRTGMNDLNAKLPGNISLNVCECRNSKQHGKDSKTDHMLRLPKA 420  
 DB 361 GGEVRLHFQEGGESRTGMNDLNAKLPGNISLNVCECRNSKQHGKDSKTDHMLRLPKA 420  
 QY 421 EDRRKEQWETKHQTERKIPKYVPPHLSPKKWLGTPIEMRRMPCGIRLPILRPSANH 480  
 DB 421 EDRRKEQWETKHQTERKIPKYVPPHLSPKKWLGTPIEMRRMPCGIRLPILRPSANH 480  
 QY 481 TVTIRVDLLRAGEVPKPPHYKDLWNKVKMPCSEONLYPVEDENGERTAGSRWELIQ 540  
 DB 481 TVTIRVDLLRAGEVPKPPHYKDLWNKVKMPCSEONLYPVEDENGERTAGSRWELIQ 540  
 QY 541 TALLNKFTFQNLKDALIKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLIYQSILPDMVKI 600  
 DB 541 TALLNKFTFQNLKDALIKYNVAYSKKWDFTALIDFWDKVLLEAEQAHLIYQSILPDMVKI 600  
 QY 601 ALCLPNTCTQPIPLKQKMHISITMSQEOIASILLANAFCTFPRNNAKMSSEYSSYPDI 660  
 DB 601 ALCLPNTCTQPIPLKQKMHISITMSQEOIASILLANAFCTFPRNNAKMSSEYSSYPDI 660  
 QY 661 FNRLFEGRSRKPCKLKTLCYFRVTEKPTGLVTTROSLEDFFPWERCEKELTLHV 720  
 DB 661 FNRLFEGRSRKPCKLKTLCYFRVTEKPTGLVTTROSLEDFFPWERCEKELTLHV 720  
 QY 721 TYEGTIEENGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNEC 780  
 DB 721 TYEGTIEENGQGLQVDFANRFVGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNEC 780  
 QY 781 LIITGTQYSEYTGATYRWSRSHEDGSDDCERCTEIVADALHFRYLDQFVPEK 840  
 DB 781 LIITGTQYSEYTGATYRWSRSHEDGSDDCERCTEIVADALHFRYLDQFVPEK 840  
 QY 841 MRRELKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVYF 900  
 DB 841 MRRELKAYCGFLRPGVSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVYF 900  
 QY 901 TFGSELMDIYSMHIFLTERKLTVDGVYKLLLRYNBECNCSTPGPDIKLYPFIYHAY 960  
 DB 901 TFGSELMDIYSMHIFLTERKLTVDGVYKLLLRYNBECNCSTPGPDIKLYPFIYHAY 960  
 QY 961 ESCAETHDSQORTCT 976  
 DB 961 ESCAETHDSQORTCT 976

RESULT 3  
 AAU76012  
 ID AAU76012 standard; protein; 976 AA.  
 XX

AC AAU76012;  
 XX 08-MAY-2002 (first entry)  
 XX Human poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 DE Human; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX Homo sapiens.  
 XX US6333148-B1.  
 XX 25-DEC-2001.  
 XX 30-APR-1999; 99US-00302812.  
 XX 01-MAY-1998; 98US-0083768P.  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14494.  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX Claim 3; Col 55-60; 80pp; English.  
 XX The present invention relates to a new method for screening compounds for  
 XX ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 XX diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 XX inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 XX are used to treat or prevent any condition associated with DNA damage  
 XX e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 XX vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 XX Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 XX Compounds identified by the new method are more effective than known  
 XX inhibitors and have fewer side effects. The present amino acid sequence  
 XX represents the human PARG protein of the invention. This protein is one  
 XX of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 XX invention  
 XX SQ Sequence 976 AA;  
 Query Match 100.0%; Score 5190; DB 5; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNAGPGCEPCTKATRWGAATTSAAADARSFPQRORVLPDKAHVQFRVPPSPACVPG 60  
 DB 1 MNAGPGCEPCTKATRWGAATTSAAADARSFPQRORVLPDKAHVQFRVPPSPACVPG 60  
 QY 61 QAGHGRGATSLVFKQKTIISWMDTKGKTAESESLDSKENNTRIESMSSVQKDNFYQ 120  
 DB 61 QAGHGRGATSLVFKQKTIISWMDTKGKTAESESLDSKENNTRIESMSSVQKDNFYQ 120  
 QY 121 HNVKLVNVSQSLDKSLTEKSTQYLNQHOTAAAMCKWQNEGKTEQLLESEPQTIVLVE 180  
 DB 121 HNVKLVNVSQSLDKSLTEKSTQYLNQHOTAAAMCKWQNEGKTEQLLESEPQTIVLVE 180  
 QY 181 QFSNANIDRSPQNDHSDTSEENRDNQFLTTVKLANAKQTTEDEHAREAKSHOKCSKS 240  
 DB 181 QFSNANIDRSPQNDHSDTSEENRDNQFLTTVKLANAKQTTEDEHAREAKSHOKCSKS 240  
 QY 241 CHGEDCASCQDEIDVVPKSPISDVGSBDVGTSGKNDKLIROESCLGNSPPEKESEP 300

Db 241 CHPGDCASCQDEIDVVPKSLSDVSGSDVGTGSKNDKLIQBSCLGNSPPFEKSESEP 300  
 QY 301 ESPMDVDNSKNSCQDEADEETSPGFDQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360  
 Db 301 ESPMDVDNSKNSCQDEADEETSPGFDQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360  
 QY 361 GGEVPLHFOFEGGESRTGMNDLNAKLPNIISSLANVECRNSKHGKKDSKITDHLMLRPLKA 420  
 Db 361 GGEVPLHFOFEGGESRTGMNDLNAKLPNIISSLANVECRNSKHGKKDSKITDHLMLRPLKA 420  
 QY 421 EDRRKEQWETHQRTKIPKVPVPHLSPDKKWLGTPIEEMRMPCRGIRLPLRPSANH 480  
 Db 421 EDRRKEQWETHQRTKIPKVPVPHLSPDKKWLGTPIEEMRMPCRGIRLPLRPSANH 480  
 QY 481 TVTIRVDLLRAGEVPKPPFTHYKDLWDNKHVQMPQSEONLYPVEDENGERTAGSRWELIQ 540  
 Db 481 TVTIRVDLLRAGEVPKPPFTHYKDLWDNKHVQMPQSEONLYPVEDENGERTAGSRWELIQ 540  
 QY 541 TALINKFTRPQNLKDAILKYNVAYSKWDFTALIDFWDKVLEAEAAQHLIYQSIILPDMVKI 600  
 Db 541 TALINKFTRPQNLKDAILKYNVAYSKWDFTALIDFWDKVLEAEAAQHLIYQSIILPDMVKI 600  
 QY 601 ALCLPNICTQPIPLKQKMHSTMSQEQIASILANAFCTFPRNNAKMSYSSYPDIN 660  
 Db 601 ALCLPNICTQPIPLKQKMHSTMSQEQIASILANAFCTFPRNNAKMSYSSYPDIN 660  
 QY 661 FNRLFEGRSSRKPXKLTLCYFRVRTEKKPTGLVTPTRQSLDPEWERCCKPLTRLHV 720  
 Db 661 FNRLFEGRSSRKPXKLTLCYFRVRTEKKPTGLVTPTRQSLDPEWERCCKPLTRLHV 720  
 QY 721 TYEGTIEENGQMLQVDFANRFPVGGVTSAGLVQBEIRFLINPELLISRLFTVLDHNEC 780  
 Db 721 TYEGTIEENGQMLQVDFANRFPVGGVTSAGLVQBEIRFLINPELLISRLFTVLDHNEC 780  
 QY 781 LIITGTEQYSEVGTGAETVNRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPEK 840  
 Db 781 LIITGTEQYSEVGTGAETVNRSHEDGSDRDCERRCTEIVAIIDALHFRYLDQFVPEK 840  
 QY 841 MRRELKAYCGFLRPGVSSSENISAVATGNWGCAGFGDARLKALIQIAAAAAERDVVVF 900  
 Db 841 MRRELKAYCGFLRPGVSSSENISAVATGNWGCAGFGDARLKALIQIAAAAAERDVVVF 900  
 QY 901 TFGDSELMDIYSMHIFLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPIYHAY 960  
 Db 901 TFGDSELMDIYSMHIFLTERKLTGVDVYKLLRYNNEECRCNCTPGPDIKLYPIYHAY 960  
 QY 961 ESCAETADHSGQRTGT 976  
 Db 961 ESCAETADHSGQRTGT 976  
 RESULT 4  
 ID ABG72279 standard; protein; 976 AA.  
 XX AC ABG72279;  
 XX DT 13-MAR-2003 (first entry)  
 XX DE Human poly (ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX KW Human; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytostatic;  
 KW neuroprotective; nootropic; antiparkinsonian; cardiant; vasotrophic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX OS Homo sapiens.  
 XX PN US2002132328-A1.

XX 19-SEP-2002.  
 PD 09-OCT-2001; 2001US-00973451.  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEJ/) AME J.  
 PA (LINW/) LIN W.  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX MPI; 2003-155895/15.  
 DR N-PSDB; ABX14478.  
 XX New nucleic acid molecule encoding a polypeptide with poly (ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX Claim 28; Fig 16; 86pp; English.  
 XX The present invention relates to the isolation of poly (ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative  
 CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents human PARG enzyme  
 XX  
 SQ Sequence 976 AA;

Query Match 100.0%; Score 5190; DB 6; Length 976;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNAGPCEPCTKATRWGAATTSPAAADARSPPSRQRRVLDPKDAHVPFRVPPSPACVPG 60  
 Db 1 MNAGPCEPCTKATRWGAATTSPAAADARSPPSRQRRVLDPKDAHVPFRVPPSPACVPG 60  
 QY 61 QAGQHRGATSILVPFKQKITSMWDTKGITABESISLDSKENNNTIESMMSSVQKDNFYQ 120  
 Db 61 QAGQHRGATSILVPFKQKITSMWDTKGITABESISLDSKENNNTIESMMSSVQKDNFYQ 120  
 QY 121 HNVEKLVNVSQSLDKSLTEKSTQYLNOHQTAAMCKWQNEGHKTEQLLESEPTVTLVPE 180  
 Db 121 HNVEKLVNVSQSLDKSLTEKSTQYLNOHQTAAMCKWQNEGHKTEQLLESEPTVTLVPE 180  
 QY 181 QFSNANIDRSPQNDHSDTDSEENRDNQFLTIVKLANAKQTTEDEHAREAKSHQCKSKS 240  
 Db 181 QFSNANIDRSPQNDHSDTDSEENRDNQFLTIVKLANAKQTTEDEHAREAKSHQCKSKS 240  
 QY 241 CHPGEDCASCQDEIDVVPKSLSDVSGEDVGTGSKNDKLIQBSCLGNSPPFEKSESEP 300  
 Db 241 CHPGEDCASCQDEIDVVPKSLSDVSGEDVGTGSKNDKLIQBSCLGNSPPFEKSESEP 300  
 QY 301 ESPMDVDNSKNSCQDEADEETSPGFDQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360  
 Db 301 ESPMDVDNSKNSCQDEADEETSPGFDQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360  
 QY 361 GGEVPLHFOFEGGESRTGMNDLNAKLPNIISSLANVECRNSKHGKKDSKITDHLMLRPLKA 420

Db 361 GGEVRLHFOEGESTGMNDLNAKLPNGNISLNVCRNSKQHGKXSKITDHLMLPKA 420  
 QY 421 EDRKEQWETKHORTERKIPKYVPHLSPDKKWLGTPIEEMRRMPRCGIRLPLRLPSANH 480  
 Db 421 EDRKEQWETKHORTERKIPKYVPHLSPDKKWLGTPIEEMRRMPRCGIRLPLRLPSANH 480  
 QY 481 TVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNLYPVEDENGERTAGRWELIQ 540  
 Db 481 TVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNLYPVEDENGERTAGRWELIQ 540  
 QY 541 TALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLERAEQAHLIYOSILPDMVKI 600  
 Db 541 TALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLERAEQAHLIYOSILPDMVKI 600  
 QY 601 ALCLPNICTOPIPLLKQNMHSITMSQEQIASILANAFCTFFRRNAKMKSEYSSYPDIN 660  
 Db 601 ALCLPNICTOPIPLLKQNMHSITMSQEQIASILANAFCTFFRRNAKMKSEYSSYPDIN 660  
 QY 661 FNLFEGRSRKPEKLTLCYFRRVTEKPTGLVTTROSLEDPFPPWERCERKPLTRLHV 720  
 Db 661 FNLFEGRSRKPEKLTLCYFRRVTEKPTGLVTTROSLEDPFPPWERCERKPLTRLHV 720  
 QY 721 TYEGTIEENGQGLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNEC 780  
 Db 721 TYEGTIEENGQGLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLDHNEC 780  
 QY 781 LIITGTQEYSEYTGAEITYNWSHEDSGERDCCERRCTEIVADLHFRFYLDQFVPEK 840  
 Db 781 LIITGTQEYSEYTGAEITYNWSHEDSGERDCCERRCTEIVADLHFRFYLDQFVPEK 840  
 QY 841 MRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 900  
 Db 841 MRRELKAYCGFLRPGVSSENLSAVATGNWCGAFGGDARLKALIQILAAAAAERDVVVF 900  
 QY 901 TFGDSELMRDIYSNHIPLTERKLTGVDVYKLLIRYNEECRNCSTPGPDIKLYPFIYHAV 960  
 Db 901 TFGDSELMRDIYSNHIPLTERKLTGVDVYKLLIRYNEECRNCSTPGPDIKLYPFIYHAV 960  
 QY 961 ESCAETADHSGQRTGT 976  
 Db 961 ESCAETADHSGQRTGT 976  
 RESULT 5  
 AAE25629  
 ID AAE25629 standard; protein; 977 AA.  
 AC AAE25629;  
 XX  
 DT 04-NOV-2002 (first entry)  
 XX  
 DE Bovine poly adenosine diphosphate-ribose glycohydrolase (bPARG).  
 XX  
 KW Bovine; poly adenosine diphosphate-ribose glycohydrolase; bPARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotropic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytotatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 XX antisense therapy.  
 OS  
 XX Bos taurus.  
 XX US6395543-B1.  
 XX  
 PD 28-MAY-2002.  
 XX  
 PF 23-FEB-2000; 2000US-00511507.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (KENT ) UNIV KENTUCKY RES FOUND.

XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42081.  
 DR  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 PS Claim 3; Col 47-45; 77pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP ribose polymer. The  
 CC invention is useful as probes and primer molecules that can be used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is bPARG  
 XX  
 SQ Sequence 977 AA;  
 Query Match 89.3%; Score 4635.5; DB 5; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 8/4; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFSPQRRLVDPKDAHVFQRPVSSPACVP 59  
 Db 1 MSAGPGCEPCTKPRWDAAATSPPAASDARSFSPQRRLVDPKDAHVFQRPVSSSGCAL 60  
 QY 60 GQAGQHRGATSIVFQKTIISWMDTKGKTAESESLDSKNNNTRESNMSSVQKDNFY 119  
 Db 61 GRAGQHRGATSIVFQKTIISWMDTKGKTAESESLDSKNNNTRESNMSSVQKDNFY 120  
 QY 120 QHNVEKLVNYSLSLDSKLTSTQVLYNOHOTAMCKWQNEGHKTEOLLESEPTVTLVP 179  
 Db 121 QHNVEKLVNYSLSLDSKLTSTQVLYNOHOTAMCKWQNEGHKTEOLLESEPTVTLVP 180  
 QY 180 EQFSNANIDRSPQNDHSDTSEENRDNQOFLTVVKLANAKQTDEHAEAKSHQCKSK 239  
 Db 181 EQFSNANVDQSSPKDDHSDTSEENRDNQOFLTVVKLANAKQTDEHAEAKSHQCKSK 240  
 QY 240 SCHPGEDCASCQODEIDVVPKSPISDVSGEDVGTGSKNDKLIROESCLNSPFEKSE 299  
 Db 241 ACPAEACACQOQEEITDVVSESLSDTSGEDVGTGLKNANRLNRQESLNSPFEKSE 300  
 QY 300 PESPMVDVNSKNSCQDSEADETSPGFDEQDSSQTANKPSRFQARDADIEFRKRYST 359  
 Db 301 PESPMVDVNSKNSCQDSEADETSPGFDEQDSSQTANKPSRFQARDADIEFRKRYST 360  
 QY 360 KGEVRLHFOEGESTGMNDLNAKLPNGNISLNVCRNSKQHGKXSKITDHLMLPK 419  
 Db 361 KGEVRLHFOEGESTGMNDVNAKPGSTSLNVCRNSKQHGKXSKITDHLMLPK 420  
 QY 420 AEDRRKEQWETKHORTERKIPKYVPHLSPDKKWLGTPIEEMRRMPRCGIRLPLRLPSAN 479  
 Db 421 AEDRRKEQWETKHORTERKIPKYVPHLSPDKKWLGTPIEEMRRMPRCGIRLPLRLPSAN 480  
 QY 480 HTVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNLYPVEDENGERTAGRWELI 539  
 Db 481 HTVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQNLYPVEDENGERTAGRWELI 540  
 QY 540 QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLERAEQAHLIYOSILPDMVK 599  
 Db 541 QTALLNKFTRPQNLKDALIKYNVAYSKKWDFTALIDFMDKVLERAEQAHLIYOSILPDMVK 600





QY 660 NFNRLEGRSSRRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRH 719  
 DB 661 NFNRLEGRSSRRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRH 720  
 QY 720 VTVEGTIEENGQGLQVDFANRFGVGGVTSAGLVQSEIRFLINPELLISRLFTFVLDHNE 779  
 DB 721 VTVEGTIEENGQGLQVDFANRFGVGGVTSAGLVQSEIRFLINPELLISRLFTFVLDHNE 780  
 QY 780 CLIIITGTEQSYTYGAETRYWGRSHEDGSDCERRCTTEIIVADALHFRYLDQFVPE 839  
 DB 781 CLIIITGTEQSYTYGAETRYWGRSHEDGSDCERRCTTEIIVADALHFRYLDQFVPE 840  
 QY 840 KMRRLNWAYCGFLRPGVSSSEN.SAVATGNWGCAGFGGDARLKALIQILAAAAARDVY 899  
 DB 841 KIRRLNWAYCGFLRPGVSSSEN.SAVATGNWGCAGFGGDARLKALIQILAAAAARDVY 900  
 QY 900 FTFGDSELMEDIYSMHIFLTERKLTGVDVYKLLRYNNECRNCSTPGPDIKLYPFIYA 959  
 DB 901 FTFGDSELMEDIYSMHIFLTERKLTGVDVYKLLRYNNECRNCSTPGPDIKLYPFIYA 960  
 QY 960 VESCAETADHSGORTG 975  
 DB 961 VESCTQTNPQORTG 976  
 RESULT 7  
 AAU75799  
 ID AAU75799 standard; protein; 977 AA.  
 XX  
 AC AAU75799;  
 XX  
 DT 08-MAY-2002 (first entry)  
 XX  
 DE Bovine poly(ADP-ribose) glycohydrolase (PARG) protein sequence.  
 XX  
 KW Cow; poly(ADP-ribose) glycohydrolase; PARG; PARG modulator; ADP-ribose;  
 KW adenosine diphosphate-ribose; DNA repair; apoptosis; neoplasia; cancer;  
 KW inherited genetic disease; myocardial infarction; vascular stroke; aging;  
 KW neurodegeneration; Huntington's disease; Parkinson's disease;  
 KW Alzheimer's disease; neurotoxicity.  
 XX  
 OS Bos taurus.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 601..617  
 FT /note= "Represents PARG oligopeptide #1"  
 FT Region 761..770  
 FT /note= "Represents PARG oligopeptide #2"  
 FT Region 771..801  
 FT /note= "Represents PARG oligopeptide #3"  
 FT Region 849..880  
 FT /note= "Represents PARG oligopeptide #4"  
 XX  
 US6333148-B1.  
 XX  
 XX  
 XX 25-DEC-2001.  
 XX  
 XX 30-APR-1999; 99US-00302812.  
 XX  
 XX 01-MAY-1998; 98US-0083768P.  
 XX  
 XX (KENT ) UNIV KENTUCKY RES FOUND.  
 XX  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 XX WPI; 2002-153820/20.  
 XX N-PSDB; ABK14493.  
 XX  
 XX Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,  
 XX useful potentially for treating diseases associated with DNA damage, e.g.  
 XX cancer.  
 XX  
 XX Claim 3; Col 45-52; 80pp; English.  
 PS

XX The present invention relates to a new method for screening compounds for  
 CC ability to modulate activity of an enzyme that hydrolyses ADP (adenosine  
 CC diphosphate)-ribose from an ADP-ribose polymer. The compounds are  
 CC inhibitors or activators of PARG (poly(ADP-ribose) glycohydrolase) and  
 CC are used to treat or prevent any condition associated with DNA damage  
 CC e.g. neoplasia, inherited genetic diseases, myocardial infarction,  
 CC vascular stroke, aging and neurodegeneration e.g. Huntington's,  
 CC Parkinson's or Alzheimer's diseases, or neurotoxicity generally.  
 CC Compounds identified by the new method are more effective than known  
 CC inhibitors and have fewer side effects. The present amino acid sequence  
 CC represents the bovine PARG protein of the invention. This protein is one  
 CC of several PARG proteins (AAU75799 and AAU76012-AAU76015) of the  
 CC invention  
 XX  
 SQ Sequence 977 AA;  
 Query Match 89.3%; Score 4635.5; DB 5; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MNAGPCCEPCTKATRWCAATTS-PAASDARSFPRORRVLDPKDAHVQFRVPPSSPACVP 59  
 DB 1 MSAGPCCEPCTKPRWDAATSPPAASDARSFPRORRVLDSKDAFVQFRVPPSSSGCAL 60  
 QY 60 GQAGQHRGSATSLVFQKQTTTSMWDTGKIKTAESELSKNNNTREESMMSVQKDNFY 119  
 DB 61 GRAGQHRGSATSLVFQKQTTTSMWDTGKIKTAESELSKNNNTREESMMSVQKDNFY 120  
 QY 120 QHNVEKLVNVSQSLDSLTETKSTOYLNOHOTAAKCKWQNEGHTEQLLESEPTVTLVP 179  
 DB 121 QHNMEKLVNVSQSLDSLTETKSTOYLNOHOTAAKCKWQNEGHTEQLLESEPTVTLVP 180  
 QY 180 EFSNANIDRESQNDHSDTDSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQCKSK 239  
 DB 181 EQFSNANVDQSSPKDDHSDTSEESRDNQOFLTTHVLANAKQTMEDEQGEARSHQCKGK 240  
 QY 240 SCHPGSDCASQODEIDVVPKPLSDVGSDEVTGSKDNKILRQESCLGNSPFPEKESE 299  
 DB 241 ACHPAEACAGCQOETDVVSEPLSDTGSDEVTGSKDNKILRQESCLGNSPFPEKESE 300  
 QY 300 PESPMVDNKSNCODSEADETSPGDEQDSSSTANKPGRFOARDADIEFRKRYST 359  
 DB 301 PESPMVDNKSNCQDSEADETSPGDEQDSSSTANKPGRFOARDADIEFRKRYST 360  
 QY 360 KGGEVRLHFQFEGESRTGMDLNALPGNISLNVCECRNSKQHGKDSKITDHLMLPK 419  
 DB 361 KGGEVRLHFQFEGESRTGMDLNALPGNISLNVCECRNSKQHGKDSKITDHLMLPK 420  
 QY 420 AEDRRKEQWETHQRTERKIPKVPVPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 479  
 DB 421 AEDRRKEQWETHQRTERKIPKVPVPHLSPDKKWLGTPIEEMRMPCGIRLPLRPSAN 480  
 QY 480 HTVTIRVDLLRAGEVPKPTTHYKDLWKNKHVMPQSEQNLYPVEDENGERTAGSRWELI 539  
 DB 481 HTVTIRVDLLRAGEVPKPTTHYKDLWKNKHVMPQSEQNLYPVEDENGERTAGSRWELI 540  
 QY 540 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLSEBAQHLQYQSLPDMVK 599  
 DB 541 QTALLNKFTRPQNLKDAILKYNVAYSKKWDFTALIDFWDKVLSEBAQHLQYQSLPDMVK 600  
 QY 600 IALCLNICTOPIPLIKQKNHSITMSQEQIASLANAFCTPPRNNAKWKSEYSPDI 659  
 DB 601 IALCLNICTOPIPLIKQKNHSITMSQEQIASLANAFCTPPRNNAKWKSEYSPDI 660  
 QY 660 NFNRLEGRSSRRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRH 719  
 DB 661 NFNRLEGRSSRRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDEPPEWERCCKPLTRH 720  
 QY 720 VTVEGTIEENGQGLQVDFANRFGVGGVTSAGLVQSEIRFLINPELLISRLFTFVLDHNE 779  
 DB 721 VTVEGTIEENGQGLQVDFANRFGVGGVTSAGLVQSEIRFLINPELLISRLFTFVLDHNE 780

QY 780 CLITGTEQSYTGYAETIRWSRSHEDSGSRDDCERRCTEIVADALHFRYLDQFVPE 839  
 DB 781 CLITGTEQSYTGYAETIRWSRSHEDSGSRDDCERRCTEIVADALHFRYLDQFVPE 840  
 QY 840 KMRRLNKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 899  
 DB 841 KIRRLNKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 900  
 QY 900 FTFGSELMDIYSMHIFLTERKLVGVYKLLRYNNEECNCSCTPGPDIKLYPFIYHA 959  
 DB 901 FTFGSELMDIYSMHIFLTERKLVGVYKLLRYNNEECNCSCTPGPDIKLYPFIYHA 960  
 QY 960 VESCAETADHSQRTG 975  
 DB 961 VESCTQTNNPQORTG 976  
 RESULT 8  
 ASG72278  
 ID ABG72278 standard; protein; 977 AA.  
 XX  
 AC ABG72278;  
 XX  
 DT 13-MAR-2003 (first entry)  
 XX  
 DE Bovine poly(ADP-ribose) glycohydrolase (PARG) enzyme.  
 XX  
 KW Bovine; poly(ADP-ribose) glycohydrolase; PARG; PARG expression;  
 KW cellular response; DNA damage; neoplastic disorder inducing agent;  
 KW oxidative stress; neoplastic disorder; myocardial infarction;  
 KW vascular stroke; neurodegenerative disorder; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; inborn genetic error;  
 KW reperfusion; ischaemia; aging; neurotoxicity; PARG activity; cytosolic;  
 KW neuroprotective; nontropic; antiparkinsonian; cardiant; vasotropic;  
 KW anticonvulsant; cerebroprotective; enzyme.  
 XX  
 OS Bos taurus.  
 XX  
 PN US2002132328-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 09-OCT-2001; 2001US-00973451.  
 XX  
 PR 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 PA (JACO/) JACOBSON M K.  
 PA (JACO/) JACOBSON E L.  
 PA (AMEU/) AME J.  
 PA (LINW/) LIN W.  
 XX  
 PI Jacobson MK, Jacobson EL, Ame J, Lin W;  
 XX  
 DR WPI: 2003-155895/15.  
 DR N-PSDB; ABX14477.  
 XX  
 PT New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)  
 PT glycohydrolase activity, for preventing, treating, or ameliorating a  
 PT disease condition, e.g. neoplastic disorder, myocardial infarction or  
 PT vascular stroke.  
 XX  
 PS Claim 28; Fig 16; 86pp; English.  
 XX  
 CC The present invention relates to the isolation of poly(ADP-ribose)  
 CC glycohydrolase (PARG) from several species, and the polynucleotide  
 CC sequences encoding them. Methods for inhibiting PARG expression or  
 CC overexpressing PARG are also disclosed. PARG is involved in the cellular  
 CC response to DNA damage, and is associated with the body's response to  
 CC neoplastic disorder inducing agents and oxidative stress. The  
 CC polynucleotide sequences encoding PARG and PARG modulators are useful for  
 CC preventing, treating, or ameliorating diseases such as neoplastic  
 CC disorders, myocardial infarction, vascular stroke, neurodegenerative

CC disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's  
 CC disease), inborn genetic errors, reperfusion following ischaemia, aging,  
 CC and neurotoxicity. The polynucleotide sequences are also useful in gene  
 CC therapy. The methods are useful for identifying an agent that modulates  
 CC PARG activity, identifying a mutant PARG allele in an individual, or  
 CC screening candidate molecules for PARG modulating activity. The present  
 CC sequence represents bovine PARG enzyme  
 XX  
 SQ Sequence 977 AA;  
 Query Match 89.3%; Score 4635.5; DB 6; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MNAGPGCEPCTKATRWGAATTS-PAASDARFPQRVRVLDPKDAHVQFRVPPSPACVP 59  
 DB 1 MSAGPGCEPCTKPRWDAAATSPPAASDARFPQRVRVLDPKDAHVQFRVPPSSGCGAL 60  
 QY 60 GQAGQHRGSATSLVFKQKTTTSWMDTKGKTAESSELSKNNNTTRESMSSVQKDNFY 119  
 DB 61 GRAGQHRGSATSLVFKQKTTTSWMDTKGKTVSESLSHKNNNTTRESMSSVQKDNFY 120  
 QY 120 QHNVEKLVVVSQSLSDKSLTEKSTOYLNOHOTAAWCKWQNGKHTEQLLSEPOVTTLVP 179  
 DB 121 QHNMEKLVVVSQSLSDKSLTEKSTOYLNOHOTAAWCKWQNGKHTEQLLSEPOVTTLVP 180  
 QY 180 EQFSNANIDRSPQNDHSDTDSSENRDNOQFLTTVKLANAKQTTEDEHAREAKSHQCKSK 239  
 DB 181 EQFSNANVDSQSPKDDHSDTDSSENRDNOQFLTTVKLANAKQTTEDEHAREAKSHQCKSK 240  
 QY 240 SCHPGEDCASQOQDEIDVVPKSPVSDVSGSEVGTGSKNDKLIQOESCLGNSPPPEKSE 299  
 DB 241 ACHPAEACAGCQOQDEIDVVPKSPVSDVSGSEVGTGSKNDKLIQOESCLGNSPPPEKSE 300  
 QY 300 PESPMVDVNSKNSCODSEADEETSPGDEQDSSTANKPSRFOARDADIEPRKRYST 359  
 DB 301 PESPMVDVNSKNSCODSEADEETSPGDEQDSSTANKPSRFOARDADIEPRKRYST 360  
 QY 360 KGEVRLHFQFEGGESRTGMNDLNAKLPGNTSSLNVECRNSKQHGKDKSKITDHLMLPK 419  
 DB 361 KGEVRLHFQFEGGESRAGMNDVNAKRPGSTSSLNVECRNSKQHGKDKSKITDHPMRVPK 420  
 QY 420 AEDRRKEQWETHQORTERKIPKYVPHLSPKKMLGTFIEEMRMRPGCIGLPLRPSAN 479  
 DB 421 AEDRRKEQWETHQORTERKIPKYVPHLSPKKMLGTFIEEMRMRPGCIGLPLRPSAN 480  
 QY 480 HTVTIRVDLLRAGEVPKPPFPHYKDLMDNKHVKMPCSEONLYPVDENGERTAGSRWELI 539  
 DB 481 HTVTIRVDLLRAGEVPKPPFPHYKDLMDNKHVKMPCSEONLYPVDENGERTAGSRWELI 540  
 QY 540 QTALLNKPTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLERAQAHLQSLIPDMVK 599  
 DB 541 QTALLNKPTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLERAQAHLQSLIPDMVK 600  
 QY 600 IALCLPNICTOPIPLLKQNMHSITMSQEQIASILLANAFCTFPRRNAMKSESSYPDI 659  
 DB 601 IALCLPNICTOPIPLLKQNMHSITMSQEQIASILLANAFCTFPRRNAMKSESSYPDI 660  
 QY 660 NFNRLEFGRSSRKEKLTLCFYFRRVTEKKPTGLVTFTRQSLDFEPWERCPELRLH 719  
 DB 661 NFNRLEFGRSSRKEKLTLCFYFRRVTEKKPTGLVTFTRQSLDFEPWERCPELRLH 720  
 QY 720 VTYEGTIEENGQGLQVDFANRFPVGGVTSAGLVQEBIRFLINPILINELISRLFTFVLDHNE 779  
 DB 721 VTYEGTIEENGQGLQVDFANRFPVGGVTSAGLVQEBIRFLINPILINELISRLFTFVLDHNE 780  
 QY 780 CLITGTEQSYTGYAETIRWSRSHEDSGSRDDCERRCTEIVADALHFRYLDQFVPE 839  
 DB 781 CLITGTEQSYTGYAETIRWSRSHEDSGSRDDCERRCTEIVADALHFRYLDQFVPE 840  
 QY 840 KMRRLNKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 899  
 DB 841 KIRRLNKAYCGFLRPGVSSSENLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 900

QY 900 FTFGDSELMRDIIYSMHILFETKLTWCDVYKLLRYNNECRNCSTPGPDIKLYPIIYHA 959  
 Db 901 FTFGDSELMRDIIYSMHILFETKLTWCDVYKLLRYNNECRNCSTPGPDIKLYPIIYHA 960  
 QY 960 VESCAETADHSGQRTG 975  
 Db 961 VESCTOTNQPGQRTG 976  
 RESULT 9  
 AAE25631  
 ID AAE25631 standard; protein; 968 AA.  
 XX AAE25631;  
 AC AAE25631;  
 DT 04-NOV-2002 (first entry)  
 XX Murine poly adenosine diphosphate-ribose glycohydrolase (PARG).  
 DE  
 XX Murine; poly adenosine diphosphate-ribose glycohydrolase; PARG; enzyme;  
 KW ADP-ribose; neoplastic disorder; genotoxic oxidative stress; neurotoxic;  
 KW cardiac disorder; neuronal disorder; reperfusion injury; neurotoxicity;  
 KW Alzheimer's disease; Huntington's disease; Parkinson's disease; cardiac;  
 KW cytostatic; vasotropic; neuroprotective; anticonvulsant; gene therapy;  
 KW antisense therapy.  
 XX  
 OS Mus musculus.  
 XX  
 XX US6395543-B1.  
 PN  
 XX 28-MAY-2002.  
 PD  
 XX 23-FEB-2000; 2000US-00511507.  
 PF  
 XX 01-MAY-1998; 98US-0083768P.  
 PR 30-APR-1999; 99US-00302812.  
 XX  
 XX (KENT) UNIV KENTUCKY RES FOUND.  
 PA  
 XX Jacobson MK, Jacobson EL, Ame J, Lin W;  
 PI WPI; 2002-535641/57.  
 XX N-PSDB; AAD42083.  
 DR  
 XX New nucleic acid molecule encoding bovine poly adenosine diphosphate-  
 PT ribose glycohydrolase involved in cellular response to DNA damage,  
 PT inhibition of which is useful for treating neoplastic disorders and  
 PT neurodegenerative diseases.  
 XX  
 XX Claim 3; Col 63-68; 77pp; English.  
 PS  
 XX The invention relates to an isolated nucleic acid molecule which encodes  
 CC a bovine poly (adenosine diphosphate (ADP)-ribose) glycohydrolase (bPARG)  
 CC which catalyses release of ADP-ribose from an ADP-ribose polymer. The  
 CC invention is useful as probes and primer molecules that can used in  
 CC hybridisation assays and polymerase chain reaction (PCR) amplification.  
 CC The knowledge of the nucleotide sequence of the PARG gene permits the  
 CC preparation of antisense therapeutics containing sequences complementary  
 CC to the mRNA of PARG gene. The antisense therapeutic are useful to treat  
 CC neoplastic disorders and conditions caused by genotoxic oxidative stress  
 CC e.g., cardiac disorders, neuronal disorders, reperfusion injury,  
 CC neurotoxicity, Alzheimer's disease, Huntington's disease and Parkinson's  
 CC disease. The invention is useful in gene therapy and antisense therapy.  
 CC The present sequence is murine PARG  
 XX  
 XX Sequence 968 AA;

Query Match 84.7%; Score 4395; DB 5; Length 968;  
 Best Local Similarity 85.7%; Pred. No. 0;  
 Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
 QY 1 MNAGPCEPCTKATRWGAA-TTSPAASDARSFPRQRRLDPKDAHVQFVRPSSPACVP 59

Db 1 MSAGPCEPCTKATRWGAA-TTSPAASDARSFPRQRRLDPKDAHVQFVRPSSPACVP 59  
 QY 60 GQAGQHRGATSIVFKQKTIITSMMDTKIGIKTAESLSDSKENNNTRIESMMSSVQKDNFY 119  
 Db 60 GRAGPERGNATSFVFKQKTIITSMMDTKIGIKTAESLSDSKENNNTRIESMMSSVQKDNFY 116  
 QY 120 QHNVEKLVNVSQSLSDKSLTEKSTQVILNQHOTAMCKWQNEGHKHTQLLESEBQTTLVLP 179  
 Db 117 PHKVEKLVNVPQLNLDKSPTEKSSQVILNQHOTAMCKWQNEGHKHTQLLESEBQTTLVLP 176  
 QY 180 EQFSNANIDRSPQNDHSDTDSENRDNOQFLTTVKLANAKQTTEDHEAREAKSHOKCSK 239  
 Db 177 KQLSNANIGQSPHTDDHSDTDSENRDNOQFLTPIKLANTKPTVGGQAR---SNCKCSG 233  
 QY 240 SCHPGEDCASCOQDEIDVVPKSPLSVGSSEDTGSGKNDNKLIRQBSCLGNSPPFKESE 299  
 Db 234 SRQSVKDCGTCQEQEVDVLPESPLSDVGAEDIGTGPKNDNKLGTQESSLGDSPPFKESE 293  
 QY 300 PESPMVDNKSNSQDSEADETSFGDEQDSSSQSTANKPSRFOARDADIEFRKRYST 359  
 Db 294 PESPMVDNRSNSQDSEADETSFVDEQDD-RSSQTANKLSSCCQAREADGDLRKRYLT 352  
 QY 360 KGEVRLHFOFEGESRTGWNLDNAKLPGNISLNVCEKNSKQHGKDKSITDHLMLPLK 419  
 Db 353 KGEVRLHFOFEGESRTGWNLDNAKLPGNISLNVCEKNSKQHGKDKSITDHLMLPLK 411  
 QY 420 AEDRKEQWETKHQRTKIPKYPVPHLSDPDKWGLTPIEEMRMPCGIRLPLRPSAN 479  
 Db 412 SEDRRKEQCEVHRQRTKIPKYPVPHLSDPDKWGLTPIEEMRMPCGIRLPLRPSAN 471  
 QY 480 HTVTIRVLLRAGEVPKPPPTHYKOLWKNKHVMPCEQNLVPEDENGERTAGSWELI 539  
 Db 472 HTVTIRVLLRAGEVPKPPPTHYKOLWKNKHVMPCEQNLVPEDENGERTAGSWELI 531  
 QY 540 QTALLNKFTPNQDKAILKYNVAYSKKWDFDALIDFWDKVLSEAAQHLVQSLPDMVK 599  
 Db 532 QTALLNKFTPNQDKAILKYNVAYSKKWDFDALIDFWDKVLSEAAQHLVQSLPDMVK 591  
 QY 600 IALCLPNICTQPIPLLKQKNSHTVMSQSIASLLANAFCTFPRNNAKMKSEYSPDI 659  
 Db 592 IALCLPNICTQPIPLLKQKNSHTVMSQSIASLLANAFCTFPRNNAKMKSEYSPDI 651  
 QY 660 NNRLPEGRSSRKPKLTKLFCYFRVTEKKTGLVTFTRQSLDPPPEWERCEKPLTRLH 719  
 Db 652 NNRLPEGRSSRKPKLTKLFCYFRVTEKKTGLVTFTRQSLDPPPEWERCEKPLTRLH 711  
 QY 720 VTYEGTIEENGQGLQVDFPANRFVGGVTSAGLVOEIRFLINPELIIISRLFTEVLHDNE 779  
 Db 712 VTYEGTIEENGQGLQVDFPANRFVGGVTSAGLVOEIRFLINPELIIISRLFTEVLHDNE 771  
 QY 780 CLIIITGTOYSYTYAETFRWRSRSHEDGSDRDCERRCTEIVADALHFRYRLQFVPE 839  
 Db 772 CLIIITGTOYSYTYAETFRWRSRSHEDGSDRDCERRCTEIVADALHFRYRLQFVPE 831  
 QY 840 KMRRLNAYCGFLRPGVSSNLSAVATGNMCGGAFGGDARLALIKIILAAAAARDVY 899  
 Db 832 KMRRLNAYCGFLRPGVSSNLSAVATGNMCGGAFGGDARLALIKIILAAAAARDVY 891  
 QY 900 FTFGDSELMRDIIYSMHILFETKLTWCDVYKLLRYNNECRNCSTPGPDIKLYPIIYHA 959  
 Db 892 FTFGDSELMRDIIYSMHILFETKLTWCDVYKLLRYNNECRNCSTPGPDIKLYPIIYHA 951  
 QY 960 VESCAETADHSGQRTG 976  
 Db 952 VESCAETADHSGQRTG 976

RESULT 10  
 AAU76022 standard; protein; 968 AA.  
 XX  
 AC AAU76022;



XX	25-DEC-2001.	
PD		
XX		
XX	30-APR-1999;	99US-00302812.
PF		
XX		
XX	01-MAY-1998;	98US-0083768P.
PR		
XX		
XX	(KENT ) UNIV KENTUCKY RES FOUND.	
PA		
XX		
XX	Jacobson MK, Jacobson EL, Ame J, Lin W;	
PI		
XX	WPI; 2002-153820/20.	
XX	N-PSDB; ABK14495.	
DR		
XX		
XX	Screening compounds for modulation of poly(ADP-ribose) glycohydrolase,	
PT	useful potentially for treating diseases associated with DNA damage, e.g.	
PT	cancer.	
XX		
XX	Claim 3; Col 63-68; 80pp; English.	
PS		
XX		
XX	The present invention relates to a new method for screening compounds for	
CC	ability to modulate activity of an enzyme that hydrolyses ADP (adenosine	
CC	diphosphate)-ribose from an ADP-ribose polymer. The compounds are	
CC	inhibitors or activators of PARP (poly(ADP-ribose) glycohydrolase) and	
CC	are used to treat or prevent any condition associated with DNA damage	
CC	e.g. neoplasia, inherited genetic diseases, myocardial infarction,	
CC	vascular stroke, aging and neurodegeneration e.g. Huntington's,	
CC	Parkinson's or Alzheimer's diseases, or neurotoxicity generally.	
CC	Compounds identified by the new method are more effective than known	
CC	inhibitors and have fewer side effects. The present amino acid sequence	
CC	represents the mouse PARP protein of the invention. This protein is one	
CC	of several PARP proteins (AAU75799 and AAU76012-AAU76015) of the	
CC	invention	
XX		
XX	Sequence 968 AA;	
SQ		

Db	472	HTVTRVVDLLRAGEVPPFPFTHYKDLNDKHKVKNPCSEQNLYPVDENGERTAGSRWELI	531
Qy	540	QTALLNKFTFQNLKDAILKYNVAYSKKDFETALIDFWDKVLEAEAAQHLVQSILPDMVK	599
Db	532	QTALLNKFTFQNLKDAILKYNVAYSKKDFETALVDFWDKVLAEAAQHLVQSILPDMVK	591
Qy	600	IALCLPNICTQPIPLLKQKNHSTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDI	659
Db	592	IALCLPNICTQPIPLLKQKNHSTMSQEQIASILLANAFCTFPRRNAMKSEYSSYPDI	651
Qy	660	NFNRLFGRRSRKPEKLTLCFCFRRVTEKKPTGLVTFTRQSLDFPEWERCPELTRLH	719
Db	652	NFNRLFGRRSRKPEKLTLCFCFRRVTEKKPTGLVTFTRQSLDFPEWERCPELTRLH	711
Qy	720	VTYEGTTEENGQGMQLQVDFANRFVGGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNE	779
Db	712	VTYEGTTEENGQGMQLQVDFANRFVGGGVTCAGLVQEEIRFLINPELIVSRFLTVDHNE	771
Qy	780	CLIIITGTEQSEYTGAEYTRWRSRHSDGSGRDCRCRCTEIVADALHFRYLDQFVPE	839
Db	772	CLIIITGTEQSEYTGAEYTRWRSRHSDGSGRDCRCRCTEIVADALHFRYLDQFVPE	831
Qy	840	KMRRELKAYCGFLRPGVSSENISAVATGNWCGAFGQDARLKALIQILAAAAAERDVVY	899
Db	832	KVRRELKAYCGFLRPGVPSENISAVATGNWCGAFGQDARLKALIQILAAAAAERDVVY	891
Qy	900	FTFGDSLMDIYSMHITFLTERKLTUGDVYKLLIRYNEECRCNCSTPGPDIKLFPFIYHA	959
Db	892	FTFGDSLMDIYSMHITFLTERKLDVGKVKLLIRYNEECRCNCSTPGPDIKLFPFIYHA	951
Qy	960	VESCAETADHSQORTGT 976	
Db	952	VESSAETDMPGQXGT 968	
RESULT 12			
ABG72280			
ID	ABG72280 standard; protein; 968 AA.		
XX			
AC	ABG72280;		
XX			
DT	13-MAR-2003 (first entry)		
XX			
DE	Mmurine poly (ADP-ribose) glycohydrolase (PARG) enzyme.		
XX			
KW	Marine; poly (ADP-ribose) glycohydrolase; PARG; PARG expression;		
KW	cellular response; DNA damage; neoplastic disorder inducing agent;		
KW	oxidative stress; neoplastic disorder; myocardial infarction;		
KW	vascular stroke; neurodegenerative disorder; Alzheimer's disease;		
KW	Parkinson's disease; Huntington's disease; inborn genetic error;		
KW	reperfusion; ischaemia; aging; neurotoxicity; PARG activity; mouse;		
KW	cytostatic; neuroprotective; neurotropic; antiparkinsonian; cardiant;		
KW	vasotropic; anticonvulsant; cerebroprotective; enzyme.		
OS	Mus musculus.		
XX			
PN	US2002132328-A1.		
XX			
PD	19-SEP-2002.		
XX			
PF	09-OCT-2001; 2001US-00973451.		
XX			
PR	01-MAY-1998; 98US-0083768P.		
PR	30-APR-1999; 99US-00302812.		
XX			
PA	(JACO/) JACOBSON M K.		
PA	(JACO/) JACOBSON E L.		
PA	(AMEJ/) AME J.		
PA	(LINW/) LIN W.		
XX			
PI	Jacobson MK, Jacobson EL, Ame J, Lin W;		
XX			

DR	WPI; 2003-155895/15.	QY	600	TALCLPNICTOPIPLKOKMHSITMSQEOIASILANAFCTTFRNNAKWKSEYSSYPDI	659
DR	N-PSDB; ABX14479.	Db	592	TALCLPNICTOPIPLKOKMHSITMSQEOIASILANAFCTTFRNNAKWKSEYSSYPDI	651
XX					
PT	New nucleic acid molecule encoding a polypeptide with poly(ADP-ribose)	QY	660	NFNRLFEGRSSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTRLH	719
PT	Glycylhydrolase activity, for preventing, treating, or ameliorating a	Db	652	NFNRLFEGRSSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDFPEWERCEKPLTRLH	711
PT	disease condition, e.g. neoplastic disorder, myocardial infarction or				
PT	vascular stroke.				
XX					
PS	Claim 28; Fig 16; 86pp; English.	QY	720	VTEGTIENGQMLQVDFANRFGVGGVTSAGLVOBEIRFLINPELIISRLTEVLDHNE	779
XX		Db	712	VTEGTIENGQMLQVDFANRFGVGGVTSAGLVOBEIRFLINPELIISRLTEVLDHNE	771
CC	The present invention relates to the isolation of poly(ADP-ribose)	QY	780	CLIIITGTQYSEYTGAEYRWMSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPE	839
CC	glycylhydrolase (PARG) from several species, and the polynucleotide	Db	772	CLIIITGTQYSEYTGAEYRWMSRSHEDGSDRDCERRCTEIVADALHFRYLDQFVPE	831
CC	sequences encoding them. Methods for inhibiting PARG expression or				
CC	overexpressing PARG are also disclosed. PARG is involved in the cellular				
CC	response to DNA damage, and is associated with the body's response to				
CC	neoplastic disorder inducing agents and oxidative stress. The				
CC	polynucleotide sequences encoding PARG and PARG modulators are useful for	QY	840	KMRRLNKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAARDVVY	899
CC	preventing, treating, or ameliorating diseases such as neoplastic	Db	832	KVRRLNKAYCGFLRPGVSSNLSAVATGNWCGAFGGDARLKALIQILAAAAARDVVY	891
CC	disorders, myocardial infarction, vascular stroke, neurodegenerative				
CC	disorders (e.g. Alzheimer's disease, Parkinson's disease or Huntington's	QY	900	FTFGDSELMRDYSHHILFTRKLTGVDVYKLLIYVNEECNCSTPGPDIKLYPIYHA	959
CC	disease), inborn genetic errors, reperfusion following ischaemia, aging,	Db	892	FTFGDSELMRDYSHHILFTRKLTGVDVYKLLIYVNEECNCSTPGPDIKLYPIYHA	951
CC	and neurotoxicity. The polynucleotide sequences are also useful in gene				
CC	therapy. The methods are useful for identifying an agent that modulates				
CC	PARG activity, identifying a mutant PARG allele in an individual, or	QY	960	VESCAETADHSGQRTGT 976	
CC	screening candidate molecules for PARG modulating activity. The present	Db	952	VESSAETDMPGQKAGT 968	
CC	sequence represents murine PARG enzyme				
XX					
SQ	Sequence 968 AA;				
Query Match 84.7%; Score 4395; DB 6; Length 968;					
Beat Local Similarity 85.7%; Pred. No. 0;					
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;					
QY	1 MNAGCPCRCYATWGA--TTSPAASDARSPPSRQRRVLDPKDAHQVQRPVPPSPACVP 59	XX	AC	ABBS9491;	
Db	1 MSAGPGWEPCTKA--RWGAAGTSAPTASRSRPFGEQRVLDPKDAVPQRPVPPSPACVS 59	XX	DT	26-MAR-2002 (first entry)	
QY	60 GQAGQHRGSATSLVFQKTIITSMWDTKGIKTAESELSKNNNTRISMSVQKDNFY 119	DE	DE	Drosophila melanogaster polypeptide SEQ ID NO 5265.	
Db	60 GRAGPHRGNATSFVFKQKTIITWMTKGPKEIAE---SKNNNTRISMSVQKDNFY 116	XX	KW	Drosophila; developmental biology; cell signalling; insecticide;	
QY	120 QHNVKLVNVSQSLDKSLTEKSTOYLNOHOTAAMCKWQNECKHTEQLLESPTVTLVP 179	XX	OS	Drosophila melanogaster.	
Db	117 PHKVEKLENVQNLNDKPTKSSQYLNQOQTASVCKWQNECKHAEQLLASEPPAGTLP 176	XX	PN	WO200171042-A2.	
QY	180 RQFSNANTDRSPQNDHSDTSEENRDNQOPLTTVKLANAKOTTEDEHAREAKSHQKSK 239	XX	PD	27-SEP-2001.	
Db	177 KQLSNANIGQSPHTDDHSDTDEHEDRDNQQLTPTKLANTKPTVGDGQAR---SNCKCSG 233	XX	PF	23-MAR-2001; 2001WO-US009231.	
QY	240 SCHPGEDCASCQOEDIVVPSPLSDVGSSEVGTGSKNDNKLIRQESCLGNSPPPEKESE 299	XX	PR	23-MAR-2000; 2000US-0191637P.	
Db	234 SRQSVKDCGTCQOEEVDVLPESPLSDVGAEDIGTGPKNDKLTGQESSLGDSPPPEKESE 293	XX	PR	11-JUL-2000; 2000US-00614150.	
QY	300 PESPMVDVNSKNSQDSSEADETSPGFDEQDSSQSTANKPSRPFQARDADIEFRKRYST 359	XX	PA	(PEKE ) PE CORP NY.	
Db	294 PESPMVDVNSRNSQDSSEADETSPVFDEQDD--RSSQTANKLSSQAREADGLRKRYLT 352	XX	PI	Venter JC, Adams M, Li PWD, Myers EM;	
QY	360 KGGEVRLHFQEGGSRPTGMDNLAKLPCNTSSINVECRNSKHQCKDSKITDHLMLPK 419	XX	PI	WPI; 2001-656860/75.	
Db	353 KGSEVRLHFQEF--GENNAQTSIDNAPSGNSSLNVECRSKQHGKRDSDKITDHPMRISK 411	XX	DR	N-PSDB; ABL03594.	
QY	420 AEDRKEQWETHQTERKIPKYPVPHLSPDKKMLGTPIEMRMRPRCGIRLPLRPSAN 479	PT	PT	New isolated nucleic acid detection reagent for detecting 1000 or more	
Db	412 SEDRKEQCEVHQTERKIPKYPVPHLSPDKKMLGTPIEMRMRPRCGIRLPLRPSAS 471	PT	PT	genes from Drosophila and for elucidating cell signaling and cell-cell	
QY	480 HTVTIRVLLRAGEVPKPPPTHYKDLWNKHVMPCCSQNLVYVDENGERTAGSRWELI 539	XX	PS	interactions.	
Db	472 HTVTIRVLLRAGEVPKPPPTHYKDLWNKHVMPCCSQNLVYVDENGERTAGSRWELI 531	XX	XX	Disclosure; SEQ ID NO 5265; 21pp + Sequence Listing; English.	
QY	540 QTALLNKTRPQNLKADAILKYNVASKKWDFTALIDFDWVKLEEAQHLQSLIPDMVK 599	CC	CC	The invention relates to an isolated nucleic acid detection reagent	
Db	532 QTALLNKTRPQNLKADAILKYNVASKKWDFTALIDFDWVKLEEAQHLQSLIPDMVK 591	CC	CC	capable of detecting 1000 or more genes from Drosophila. The invention is	
		CC	CC	useful in developmental biology and in elucidating cell signalling and	
		CC	CC	cell-cell interactions in higher eukaryotes for the development of	
		CC	CC	insecticides, therapeutics and pharmaceutical drugs. The invention	
		CC	CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA	
		CC	CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-	









```

301 ESPMDVNSKNSQDSEADEETSPGFDEQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360
361 GGEVLHLPQFEGESRTGNDLNKLPNISSLNVECRNSKHGKDSKI TDHLMRLPKA 420
361 GGEVLHLPQFEGESRTGNDLNKLPNISSLNVECRNSKHGKDSKI TDHLMRLPKA 420
421 EDRRKEQWETHKHQTERKIPKVPVPHLSPDKWMLGTPIEEMRMPCRGIRLPLRPSANH 480
421 EDRRKEQWETHKHQTERKIPKVPVPHLSPDKWMLGTPIEEMRMPCRGIRLPLRPSANH 480
481 TVTIRVDLLRAGEVPKPTTHYKDLWDNKHVMPQSEQNLYPVEDENGERTAGSRWELIQ 540
481 TVTIRVDLLRAGEVPKPTTHYKDLWDNKHVMPQSEQNLYPVEDENGERTAGSRWELIQ 540
541 TALLNKFRPNQKDALIKYNVAYSKKWDFTALIDFMDKVLEEAQAHLIYQSILPDMVKI 600
541 TALLNKFRPNQKDALIKYNVAYSKKWDFTALIDFMDKVLEEAQAHLIYQSILPDMVKI 600
601 ALCLPNICTQPIPLLKQKNHSITWSQEQIASLLANAFCTFPRRNAMKSEYSSYPDIN 660
601 ALCLPNICTQPIPLLKQKNHSITWSQEQIASLLANAFCTFPRRNAMKSEYSSYPDIN 660
661 FNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDPPWECEKPLTRLHV 720
661 FNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDPPWECEKPLTRLHV 720
721 TYEGTIEENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELII SRLFTVLDHNEC 780
721 TYEGTIEENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELII SRLFTVLDHNEC 780
781 LIITGTEQYSEYTGVAETRWRSRSHEDGSDRDCERRCTEIVAI DALLHFRYLDQFVPEK 840
781 LIITGTEQYSEYTGVAETRWRSRSHEDGSDRDCERRCTEIVAI DALLHFRYLDQFVPEK 840
841 MRRELKAYCGFLRPGVSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAARDVVYF 900
841 MRRELKAYCGFLRPGVSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAARDVVYF 900
901 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPIYHAV 960
901 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPIYHAV 960
961 ESCAETADHSQORTGT 976
961 ESCAETADHSQORTGT 976

RESULT 2
US-09-511-477-4
; Sequence 4, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; CURRENT FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-511-477-4

```

```

Query Match
Best Local Similarity 100.0%; Score 5190; DB 4; Length 976;
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNAGGCPCTKATRWGAATSPAAADARSPPSRORRVLDPKDAHVQFRVPSSPACVPG 60
DB 1 MNAGGCPCTKATRWGAATSPAAADARSPPSRORRVLDPKDAHVQFRVPSSPACVPG 60
QY 61 QAGQHRGSAATSLVFKQKTTISWMDTKGIKTAESBSLDSKENNNTRIESMWSVQKDNFVQ 120
DB 61 QAGQHRGSAATSLVFKQKTTISWMDTKGIKTAESBSLDSKENNNTRIESMWSVQKDNFVQ 120
QY 121 HNVEKLNVVSQSLSDKSLTEKSTOYLNOHQTAAMCKWNEGHGHTQLLESEPTVTLVPE 180
DB 121 HNVEKLNVVSQSLSDKSLTEKSTOYLNOHQTAAMCKWNEGHGHTQLLESEPTVTLVPE 180
QY 181 QFSNANIDRSQNDHSDTSEENRDNOQFTLVKLANAKQTTDEHAREAKSHQKCSKS 240
DB 181 QFSNANIDRSQNDHSDTSEENRDNOQFTLVKLANAKQTTDEHAREAKSHQKCSKS 240
QY 241 CHPGEDCASQODEIDVVPKSPLDVGSDEVTGSKNDKLIROESCLGNPPPFKESEBP 300
DB 241 CHPGEDCASQODEIDVVPKSPLDVGSDEVTGSKNDKLIROESCLGNPPPFKESEBP 300
QY 301 ESPMDVNSKNSQDSEADEETSPGFDEQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360
DB 301 ESPMDVNSKNSQDSEADEETSPGFDEQEDGSSQTANKPSRFQARDADIEFRKRYSTK 360
QY 361 GGEVLHLPQFEGESRTGNDLNKLPNISSLNVECRNSKHGKDSKI TDHLMRLPKA 420
DB 361 GGEVLHLPQFEGESRTGNDLNKLPNISSLNVECRNSKHGKDSKI TDHLMRLPKA 420
QY 421 EDRRKEQWETHKHQTERKIPKVPVPHLSPDKWMLGTPIEEMRMPCRGIRLPLRPSANH 480
DB 421 EDRRKEQWETHKHQTERKIPKVPVPHLSPDKWMLGTPIEEMRMPCRGIRLPLRPSANH 480
QY 481 TVTIRVDLLRAGEVPKPTTHYKDLWDNKHVMPQSEQNLYPVEDENGERTAGSRWELIQ 540
DB 481 TVTIRVDLLRAGEVPKPTTHYKDLWDNKHVMPQSEQNLYPVEDENGERTAGSRWELIQ 540
QY 541 TALLNKFRPNQKDALIKYNVAYSKKWDFTALIDFMDKVLEEAQAHLIYQSILPDMVKI 600
DB 541 TALLNKFRPNQKDALIKYNVAYSKKWDFTALIDFMDKVLEEAQAHLIYQSILPDMVKI 600
QY 601 ALCLPNICTQPIPLLKQKNHSITWSQEQIASLLANAFCTFPRRNAMKSEYSSYPDIN 660
DB 601 ALCLPNICTQPIPLLKQKNHSITWSQEQIASLLANAFCTFPRRNAMKSEYSSYPDIN 660
QY 661 FNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDPPWECEKPLTRLHV 720
DB 661 FNRLFEGRSSRKPEKLTLCYFRRVTEKKPTGLVTFTRQSLDPPWECEKPLTRLHV 720
QY 721 TYEGTIEENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELII SRLFTVLDHNEC 780
DB 721 TYEGTIEENGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELII SRLFTVLDHNEC 780
QY 781 LIITGTEQYSEYTGVAETRWRSRSHEDGSDRDCERRCTEIVAI DALLHFRYLDQFVPEK 840
DB 781 LIITGTEQYSEYTGVAETRWRSRSHEDGSDRDCERRCTEIVAI DALLHFRYLDQFVPEK 840
QY 841 MRRELKAYCGFLRPGVSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAARDVVYF 900
DB 841 MRRELKAYCGFLRPGVSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAARDVVYF 900
QY 901 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPIYHAV 960
DB 901 TFGDSELMRDIYSMHIFLTERKLTVDGVYKLLRYNNEECRNCSTPGPDIKLYPIYHAV 960
QY 961 ESCAETADHSQORTGT 976
DB 961 ESCAETADHSQORTGT 976

```

RESULT 3  
US-09-511-507-4  
; Sequence 4, Application US/095111507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 4  
; LENGTH: 976  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-511-507-4

Query Match 100.0%; Score 5190; DB 4; Length 976;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNAGPGCEPCTKATRWGAATTS PAASDARSFPQRRLVDPKDAHVQFRVPPSPACVPG 60  
DB 1 MNAGPGCEPCTKATRWGAATTS PAASDARSFPQRRLVDPKDAHVQFRVPPSPACVPG 60  
QY 61 QAGHGRSATSVPFKQKTTSMWDTGKTAESLDSKENNTRIESMSSVQKONFYQ 120  
DB 61 QAGHGRSATSVPFKQKTTSMWDTGKTAESLDSKENNTRIESMSSVQKONFYQ 120  
QY 121 HNYEKLNVNLSLDSKLTSTKSTQYLNQHOHTAAMCKWQNEGKHTQELLESEPTVTLVPE 180  
DB 121 HNYEKLNVNLSLDSKLTSTKSTQYLNQHOHTAAMCKWQNEGKHTQELLESEPTVTLVPE 180  
QY 181 QFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240  
DB 181 QFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQKCSKS 240  
QY 241 CHPGEDCASCQDEIDVVPKSPISDVGSDEVTGSKNDNKLIRQESCLGNSPPFEKES 300  
DB 241 CHPGEDCASCQDEIDVVPKSPISDVGSDEVTGSKNDNKLIRQESCLGNSPPFEKES 300  
QY 301 ESPMDVNSKNSCQDSEADETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360  
DB 301 ESPMDVNSKNSCQDSEADETSPGFDEQEDGSSSQTANKPSRFQARDADIEFRKRYSTK 360  
QY 361 GGEVRLHFQEGESRTGMNDLNAKLPCLNSSLNVECRNSKHGKDKSKITDHLMLPKA 420  
DB 361 GGEVRLHFQEGESRTGMNDLNAKLPCLNSSLNVECRNSKHGKDKSKITDHLMLPKA 420  
QY 421 EDRKKEQWETKHQTERKIPKYVPPHLSPDKWLGTPIEMRMMPRCGIRLPLLRPSANH 480  
DB 421 EDRKKEQWETKHQTERKIPKYVPPHLSPDKWLGTPIEMRMMPRCGIRLPLLRPSANH 480  
QY 481 TVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQLNYPVEDENGERTAGSRWELIQ 540  
DB 481 TVTIRVDLLRAGEVPKPPFTHYKDLMDNKHVKMPCSEQLNYPVEDENGERTAGSRWELIQ 540  
QY 541 TALLNKETRPONLKDAILKNVAYSKKWDTALIDFWDKVLBEAEAHLYQSILPDMVKI 600  
DB 541 TALLNKETRPONLKDAILKNVAYSKKWDTALIDFWDKVLBEAEAHLYQSILPDMVKI 600  
QY 601 ALCLPNICTOPIPLLKQNMHSITMSQEQIASILLANAFCTEPRRNAMKSESYSPDIN 660  
DB 601 ALCLPNICTOPIPLLKQNMHSITMSQEQIASILLANAFCTEPRRNAMKSESYSPDIN 660

QY 661 FNRLFEGRSSRKDEKLTLCYFRRVTEKPTGLVTTROSLEDFPWEWERCEKPLTRLHV 720  
DB 661 FNRLFEGRSSRKDEKLTLCYFRRVTEKPTGLVTTROSLEDFPWEWERCEKPLTRLHV 720  
QY 721 TYEGTIEENGQGMLOVDFAFNRVFGGGVTSAGLVQOEIRFLINPELLIISRLFTVLDHNEC 780  
DB 721 TYEGTIEENGQGMLOVDFAFNRVFGGGVTSAGLVQOEIRFLINPELLIISRLFTVLDHNEC 780  
QY 781 LIITGTEOYSEYTGATYEWSSHEDGSDRDCERCTEIVADALHFRYLDQFVPEK 840  
DB 781 LIITGTEOYSEYTGATYEWSSHEDGSDRDCERCTEIVADALHFRYLDQFVPEK 840  
QY 841 MRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGCDARLKALIQILAAAAERDVVVF 900  
DB 841 MRRELKAYCGFLRPGVSSSENLSAVATGNWCGAFGCDARLKALIQILAAAAERDVVVF 900  
QY 901 TFGDSELMRDIIYSMHIFLTERKLTGVGVYKLLRYNNEECRNCSTPGPDIKLPFFIHAV 960  
DB 901 TFGDSELMRDIIYSMHIFLTERKLTGVGVYKLLRYNNEECRNCSTPGPDIKLPFFIHAV 960  
QY 961 ESCAETADHSGQRTGT 976  
DB 961 ESCAETADHSGQRTGT 976

RESULT 4  
US-09-302-812-2  
; Sequence 2, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/302,812B  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER APPLICATION NUMBER: 60/083,768  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-302-812-2

Query Match 89.3%; Score 4635,5; DB 4; Length 977;  
Best Local Similarity 89.5%; Pred. No. 0;  
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
QY 1 MNAGPGCEPCTKATRWGAATTS PAASDARSFPQRRLVDPKDAHVQFRVPPSPACV 59  
DB 1 MSAGPGCEPCTKPRWDAAATSPPAASDARSFPQRRLVDPKDAHVQFRVPPSPSGCAL 60  
QY 60 QGAGHGRSATSVPFKQKTTSMWDTGKTAESLDSKENNTRIESMSSVQKONFY 119  
DB 61 GRAGHGRSATSVPFKQKTTSMWDTGKTAESLDSKENNTRIESMSSVQKONFY 120  
QY 120 QHNVEKLVNVSQSLDSKLTSTKSTQYLNQHOHTAAMCKWQNEGKHTQELLESEPTVTLV 179  
DB 121 QHNVEKLVNVSQSLDSKLTSTKSTQYLNQHOHTAAMCKWQNEGKHTQELLESEPTVTLV 180  
QY 180 QFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQKCSK 239  
DB 181 QFSNANIDRSPQNDHSDTSEENRDNQOFLTTVKLANAKQTTEDEHAREAKSHQKCSK 240  
QY 240 SCHPGEDCASCQDEIDVVPKSPISDVGSDEVTGSKNDNKLIRQESCLGNSPPFEKES 299  
DB 241 ACPHACACAGCQDEIDVVPKSPISDVGSDEVTGSKNDNKLIRQESCLGNSPPFEKES 300

300 PESMDVNSKNSQDSSEDETSFGDEQEDGSSSTANKPSRFOARDADIEPRKRYST 359  
 301 PESMDVNSKNSQDSSEDETSFGDEQEDGSSSTANKPSRFOARDADIEPRKRYST 360  
 360 KGEVRLHFQEGGESRTGMNDLNAKLPNGTSSLNVECRNSKQHGKDSKITDHLMLPK 419  
 361 KGEVRLHFQEGGESRTGMNDLNAKLPNGTSSLNVECRNSKQHGKDSKITDHLMLPK 420  
 420 AEDRKEQWETHQRTKIPKYPVPHLSPPKMLGTPIEMRRMPRCGIRLPLRPSAN 479  
 421 AEDRKEQWETHQRTKIPKYPVPHLSPPKMLGTPIEMRRMPRCGIRLPLRPSAN 480  
 480 HTVTIRVDLLRAGEVPKPPHYKDLNDKHKVMPSEONLYPVEDENGERTAGSRWELI 539  
 481 HTVTIRVDLLRAGEVPKPPHYKDLNDKHKVMPSEONLYPVEDENGERTAGSRWELI 540  
 540 QTLINKTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLAEAAOHLQSYLPDMVK 599  
 541 QTLINKTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLAEAAOHLQSYLPDMVK 600  
 600 IALCLPNICTOPIPLLKOKMHSITMSQEQIASILANAFCTFPRRNAMKSEYSSYDI 659  
 601 IALCLPNICTOPIPLLKOKMHSITMSQEQIASILANAFCTFPRRNAMKSEYSSYDI 660  
 660 NFNLFEGRSSRKPEKLTLCYFRVTEKKPTGLVTTTROSLEDFFPWERCEKLTSLH 719  
 661 NFNLFEGRSSRKPEKLTLCYFRVTEKKPTGLVTTTROSLEDFFPWERCEKLTSLH 720  
 720 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVOEEIRFLINPELIIISRLFTVELDNE 779  
 721 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVOEEIRFLINPELIIISRLFTVELDNE 780  
 780 CLIIITGTOYSEYTGAEYTYRWSHEDGSDRDCRRCCTEIVADLHFRYLDQFVPE 839  
 781 CLIIITGTOYSEYTGAEYTYRWSHEDGSDRDCRRCCTEIVADLHFRYLDQFVPE 840  
 840 KMRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 899  
 841 KMRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 900  
 900 FTFGSELMDIYSMHIFLTERKLTGVGVYKLLRYNEECNCSTPGDIIKLYPIIYHA 959  
 901 FTFGSELMDIYSMHIFLTERKLTGVGVYKLLRYNEECNCSTPGDIIKLYPIIYHA 960  
 960 VESCAETADHSGQRTG 975  
 961 VESCTQTTNQPGQRTG 976

RESULT 5  
 US-09-511-477-2  
 ; Sequence 2, Application US/09511477  
 ; Patent No. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: ABE, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAID 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; PRIOR FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:

US-09-511-477-2

Query Match 89.3%; Score 4635.5; DB 4; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;

QY 1 MNAGPCEPCTKATRWGAATTS-PAASDARSFSPQRRLVDPKDAHVQFRVPPSPACVP 59  
 DB 1 MSAGPCEPCTKPRWDAAATSPAAADARSFQQRRLVDSKDPVQFRVPPSSGAL 60  
 QY 60 GOAQHGRGATSILVFKOKTITSMWDTKGIKTAESLSDSKENNNTRIESMMSSVQKNFY 119  
 DB 61 GRAGQHGRGATSILVFKOKTITSMWDTKGIKTVESLHSHKNNNTRIESMMSSVQKNFY 120  
 QY 120 QHNVEKLNVYSQLSDKSLTEKSTQYLNQHOTQAMCKWQNEGKHTEQLLESEPTVTLVP 179  
 DB 121 QHNVEKLNVYSQLGDFKSPVEKGTQYLKQHOTQAMCKWQNEGPHSERLLESEPAVTLVP 180  
 QY 180 EQFSNANIDRSPQNDHSDTSENRRNQOPLTTVKLANAKQTTEDEHARBAKSHQKCK 239  
 DB 181 EQFSNANVDQSSPKDDHSDTSENRRNQOPLTTVKLANAKQTTEDEGQREARSHQKCK 240  
 QY 240 SCHFGEDCASQOQDEIDVVPKSPLSVDYGSSEDTGSKNDNKLIRQESCLGNSPPFEKSE 299  
 DB 241 ACHPAEACAGCQOQDETDVWSEPLSDTGSSEDTGSKNDNKLIRQESCLGNSPPFEKSE 300  
 QY 300 PESMDVNSKNSQDSSEDETSFGDEQEDGSSSTANKPSRFOARDADIEPRKRYST 359  
 DB 301 PESMDVNSKNSQDSSEDETSFGDEQEDGSSSTANKPSRFOARDADIEPRKRYST 360  
 QY 360 KGEVRLHFQEGGESRTGMNDLNAKLPNGTSSLNVECRNSKQHGKDSKITDHLMLPK 419  
 DB 361 KGEVRLHFQEGGESRTGMNDLNAKLPNGTSSLNVECRNSKQHGKDSKITDHLMLPK 420  
 QY 420 AEDRKEQWETHQRTKIPKYPVPHLSPPKMLGTPIEMRRMPRCGIRLPLRPSAN 479  
 DB 421 AEDRKEQWETHQRTKIPKYPVPHLSPPKMLGTPIEMRRMPRCGIRLPLRPSAN 480  
 QY 480 HTVTIRVDLLRAGEVPKPPHYKDLNDKHKVMPSEONLYPVEDENGERTAGSRWELI 539  
 DB 481 HTVTIRVDLLRAGEVPKPPHYKDLNDKHKVMPSEONLYPVEDENGERTAGSRWELI 540  
 QY 540 QTLINKTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLAEAAOHLQSYLPDMVK 599  
 DB 541 QTLINKTRPQNLKDALIKYNVAYSKKWDFTALIDFWDKVLAEAAOHLQSYLPDMVK 600  
 QY 600 IALCLPNICTOPIPLLKOKMHSITMSQEQIASILANAFCTFPRRNAMKSEYSSYDI 659  
 DB 601 IALCLPNICTOPIPLLKOKMHSITMSQEQIASILANAFCTFPRRNAMKSEYSSYDI 660  
 QY 660 NFNLFEGRSSRKPEKLTLCYFRVTEKKPTGLVTTTROSLEDFFPWERCEKLTSLH 719  
 DB 661 NFNLFEGRSSRKPEKLTLCYFRVTEKKPTGLVTTTROSLEDFFPWERCEKLTSLH 720  
 QY 720 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVOEEIRFLINPELIIISRLFTVELDNE 779  
 DB 721 VTYEGTIEGNGQGLQVDFANRFVGGVTSAGLVOEEIRFLINPELIIISRLFTVELDNE 780  
 QY 780 CLIIITGTOYSEYTGAEYTYRWSHEDGSDRDCRRCCTEIVADLHFRYLDQFVPE 839  
 DB 781 CLIIITGTOYSEYTGAEYTYRWSHEDGSDRDCRRCCTEIVADLHFRYLDQFVPE 840  
 QY 840 KMRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 899  
 DB 841 KMRRELKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAERDVVY 900  
 QY 900 FTFGSELMDIYSMHIFLTERKLTGVGVYKLLRYNEECNCSTPGDIIKLYPIIYHA 959  
 DB 901 FTFGSELMDIYSMHIFLTERKLTGVGVYKLLRYNEECNCSTPGDIIKLYPIIYHA 960  
 QY 960 VESCAETADHSGQRTG 975  
 DB 961 VESCTQTTNQPGQRTG 976

RESULT 6  
 US-09-511-507-2  
 ; Sequence 2, Application US/09511507  
 ; Patent No. 6395543  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,507  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 2  
 ; LENGTH: 977  
 ; TYPE: PRT  
 ; ORGANISM: Bos taurus  
 ; FEATURE:  
 US-09-511-507-2

Query Match 89.3%; Score 4635.5; DB 4; Length 977;  
 Best Local Similarity 89.5%; Pred. No. 0;  
 Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
 QY 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFSPRQRRLVDPKDAHVQFRVPSSPACVP 59  
 DB 1 MSAGPGCEPCTKPRMDAAATSPPAASDARSFPGQRRLVDSKDPAPVQFRVPSSGAL 60  
 QY 60 GOAGQHRGSATSLVFQKKTITSMWDTKGITAESESDSKENNTNRTESMMSSVQKDNFY 119  
 DB 61 GRAGQHRGSATSLVFQKKTITSMWDTKGITAESESLHSENNNTRESMMSSVQKDNFY 120  
 QY 120 QHNVEKLVNYSQSLDKSLTEKSTQYLNQHTQAMCKWQEGKHTQLLESEPTVTLVP 179  
 DB 121 QHNWEKLVNYSQSLGFDKSPVEKGTQYLKQHTQAMCKWQEGHSEPLLESEPTVTLVP 180  
 QY 180 EQFSNANIDRSPQNDHSDTSDENRDNQOFLITVKLANAKQTTEDEHAREAKSHQKCSK 239  
 DB 181 EQFSNANVDQSPKDDHSDTSDENRDNQOFLITVKLANAKQTTEDEHAREAKSHQKCGK 240  
 QY 240 SCHPGEDCASCQDEIDVVPKSPLSVDGSEVDVGTGSKNDKLIROESCLGNSPPFEKSE 299  
 DB 241 ACHPAEACACQDEIDVVPKSPLSVDGSEVDVGTGSKNANLNROESCLGNSPPFEKSE 300  
 QY 300 PESPMVDNKNQSCQDSEADETSPGDEQEDGSSQTANKPSRQFQARDADIBFRKRYST 359  
 DB 301 PESPMVDNKNQSCQDSEADETSPGDEQEDGSSQTANKPSRQFQARDADITELRKSSA 360  
 QY 360 KGEVRLHFQEGESTGNDLNAKLPNMISSLVNVECRNSKQHGKQDSKITDHLRLPK 419  
 DB 361 KGEVRLHFQEGESRAGNDVNAKPGSTSLNVECRNSKQHGKQDSKITDHFMEVVK 420  
 QY 420 AEDRRKEQWTKHQRTERKIPKVPVPHLSPDKKWLGTPIEMRRMPGRLPLLRPSAN 479  
 DB 421 AEDRRKEQWTKHQRTERKIPKVPVPHLSPDKKWLGTPIEMRRMPGRLPLLRPSAN 480  
 QY 480 HTVTIRVDLLRAGEVPKPPFTHYKDLNDKNHVKMPCSEQNLVPYDENGERTAGSRWELI 539  
 DB 481 HTVTIRVDLLRAGEVPKPPFTHYKDLNDKNHVKMPCSEQNLVPYDENGERTAGSRWELI 540  
 QY 540 QTALLNKFTPQNLKDALIKTNVAYSKKWDFTALIDFWDKYLEAEAAHLYQSILPDMVK 599  
 DB 541 QTALLNKFTPQNLKDALIKTNVAYSKKWDFTALIDFWDKYLEAEAAHLYQSILPDMVK 600  
 QY 600 IALCLPNICTOPIPLLKQKNHSTMSQEQIASILANAFCTFPRRNAMKSEYSSYPDI 659

DB 601 IALCLPNICTOPIPLLKQKNHSTMSQEQIASILANAFCTFPRRNAMKSEYSSYPDI 660  
 QY 660 NFNLFEGRSSRKPEKLTLCYFRFRVTEKKPTGLVTFTRQSLDEDPFWRERCKPLRLH 719  
 DB 661 NFNLFEGRSSRKPEKLTLCYFRFRVTEKKPTGLVTFTRQSLDEDPFWRERCKPLRLH 720  
 QY 720 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIIISLFEVLQDNE 779  
 DB 721 VTYEGTIEGNGQMLQVDFANRFVGGVTSAGLVQBEIRFLINPELIIISLFEVLQDNE 780  
 QY 780 CLIIITGEQVSEYTGVAETVRWRSRSHEDGSDCERRCTEIVAIADALHFRRLVQFVPE 839  
 DB 781 CLIIITGEQVSEYTGVAETVRWRSRSHEDGSDCERRCTEIVAIADALHFRRLVQFVPE 840  
 QY 840 KMRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAERDVVY 899  
 DB 841 KMRRELKAYCGFLRPGVSSSENLSAVATGNWGCAGFGGDARLKALIQILAAAAERDVVY 900  
 QY 900 FTGDSFELMRDIYSMHIFLTERKLTGVGVYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 959  
 DB 901 FTGDSFELMRDIYSMHIFLTERKLTGVGVYKLLRYNNEECRCNCSTPGPDIKLYPFIYHA 960  
 QY 960 VESCAETADHSGQRTG 975  
 DB 961 VESCTQTNQPGQRTG 976  
 RESULT 7  
 US-09-302-812-6  
 ; Sequence 6, Application US/09302812B  
 ; Patent No. 6333148  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/302,812B  
 ; CURRENT FILING DATE: 1999-04-30  
 ; EARLIER APPLICATION NUMBER: 60/083,768  
 ; EARLIER FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-302-812-6  
 Query Match 84.7%; Score 4395; DB 4; Length 968;  
 Best Local Similarity 85.7%; Pred. No. 0;  
 Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
 QY 1 MNAGPGCEPCTKATRWGAATTS-PAASDARSFSPRQRRLVDPKDAHVQFRVPSSPACVP 59  
 DB 1 MSAGPGCEPCTKA-RWGAAGTSAFTASDSKSPFQRRLVDPKDAHVQFRVPSSPACVS 59  
 QY 60 GOAGQHRGSATSLVFQKKTITSMWDTKGITAESESLDSKENNTNRTESMMSSVQKDNFY 119  
 DB 60 GRAGPHRGATSLVFQKKTITSMWDTKGITAESESLDSKENNTNRTESMMSSVQKDNFY 116  
 QY 120 QHNVEKLVNYSQSLDKSLTEKSTQYLNQHTQAMCKWQEGKHTQLLESEPTVTLVP 179  
 DB 117 PHKVEKLVNYPQLNDKSPTEKSSQYLNQHTQAMCKWQEGKHTQLLESEPTVTLVP 176  
 QY 180 EQFSNANIDRSPQNDHSDTSDENRDNQOFLITVKLANAKQTTEDEHAREAKSHQKCSK 239  
 DB 177 KQLSNANIGQSPHTDHDSDTDEEDRDNQOFLITVKLANAKQTTEDEHAREAKSHQKCSG 233  
 QY 240 SCHPGEDCASCQDEIDVVPKSPLSVDGSEVDVGTGSKNDKLIROESCLGNSPPFEKSE 299

```

; ORGANISM: Mus musculus
; FEATURE:
US-09-511-477-6

Query Match      84.7%; Score 4395; DB 4; Length 968;
Best Local Similarity 85.7%; Pred. No. 0;
Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MNAGGCEPCTKATRWGAA--TTSAAASDARSFPSPQREVLDPKDAHVQFRVPPSPACVP 59
DB 1 NSAGFGEPCTKA-RWGAAGTSAPTASDSRFPQRRVLPKDAVQFRVPPSPACVS 59
QY 60 GOAGQHRGSATSLVFKOKTITSMWMTKIGKTAESSESLDKENNNTRIEISMSSVQKDNFY 119
DB 60 GRAGPHRGNAISFVFKQKTIITWMTKIGKTAESSE--SKENNNTRIDSMSSVQKDNFY 116
QY 120 QHNVEKLVNVSQSLDKSLTEKSTQYLNQHOHTAMCKWQNEKQKTEQLLESPQVTLVP 179
DB 117 PHKVEKLVNVSQSLDKSLTEKSTQYLNQHOHTAMCKWQNEKQKTEQLLESPQVTLVP 176
QY 180 EQFSNANIDRSPQNDHSDTSEENRDNQOFLTIVKLANAKQTTREHAREAKSHQKCSK 239
DB 177 KQLSNANIGQSPHTDDHSDTDEEDRDNQOFLTIVKLANAKQTTREHAREAKSHQKCSK 233
QY 240 SCHPGECDASCQOQDEIDVVPKSPLSVSGSEVGTGSKNDNKLIRQESCLGNSPPPEKSE 299
DB 234 SRQSVKDCCTGCOQEEVDVLPESPLSDVGAEDIGTGPKNDNKLITGQESSLGDSPPPEKSE 293
QY 300 PESPMVDNSKNSCODSEADSETSPGFBQEDGSSQSTANKPSRQARDADIEFRKRYST 359
DB 294 PESPMVDNSRNSCODSEADSETSPVFEQDD--RSSQTANKLSSCOAREADGLRKRILT 352
QY 360 KGEVRLHFQFEGGSRRTGMDLNKLPNTSSLNVECRNSKQHGKQKSKITDHLMLPK 419
DB 353 KGEVRLHFQFEGGSRRTGMDLNKLPNTSSLNVECRNSKQHGKQKSKITDHLMLPK 411
QY 420 AEDREKEQWETKHQTERKIPKYVPPHLSPPDKMLGTPIEENMRMPCGIRLPLRPSAN 479
DB 412 SEDRKEQCEVHQHQRTERKIPKYIPNLPPEKKWLTPIEENMRMPCGIRLPLRPSAS 471
QY 480 HVTIVRDLRAGEVPKPPFTHYKDLWNKHVMPCESEONLYPVEDENGERTAGSRWELI 539
DB 472 HVTIVRDLRAGEVPKPPFTHYKDLWNKHVMPCESEONLYPVEDENGERTAGSRWELI 531
QY 540 QTALLNKFTRPQNLKDAILKYNVASKKWDFTALDFWPKVLEAEAAQHLQSLILPDMVK 599
DB 532 QTALLNKFTRPQNLKDAILKYNVASKKWDFTALDFWPKVLEAEAAQHLQSLILPDMVK 591
QY 600 IALCLPNICTQPIPLLKQNNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659
DB 592 IALCLPNICTQPIPLLKQNNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651
QY 660 NFNRLEFEGSRSSRKPEKLTLCYFRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTRLH 719
DB 652 NFNRLEFEGSRSSRKPEKLTLCYFRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTRLH 711
QY 720 VTYEGTIENGQGMLOVDPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNE 779
DB 712 VTYEGTIENGQGMLOVDPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNE 771
QY 780 CLIIITEQYSEYTGVAETIRWSRSHEDGSDRDCERRCTEIVADILHFRYLDQFVPE 839
DB 772 CLIIITEQYSEYTGVAETIRWSRSHEDGSDRDCERRCTEIVADILHFRYLDQFVPE 831
QY 840 KVRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDARLKALIQIIIAAAAAERDVY 899
DB 832 KVRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDARLKALIQIIIAAAAAERDVY 891
QY 900 FTFGDSELMRDIYSMHIFLTERKLTGVGVVYKLLRYNEECRNCSTPGPDILKLYPIYHA 959
DB 892 FTFGDSELMRDIYSMHIFLTERKLTGVGVVYKLLRYNEECRNCSTPGPDILKLYPIYHA 951
QY 960 VESCAETADHSGORTGT 976
DB 952 VESCAETADHSGORTGT 976

RESULT 8
US-09-511-477-6
; Sequence 6, Application US/09511477
; Patent No. 6337202
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Blaine L.
; APPLICANT: JACOBSON, Christophe
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARC) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; FILE OF INVENTION: THEREMITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,477
; PRIOR APPLICATION NUMBER: 2000-02-23
; PRIOR FILING DATE: 09/302,812
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 6
; LENGTH: 968
; TYPE: PRT

```

```

234 SRQSVKDCCTGCOQEEVDVLPESPLSDVGAEDIGTGPKNDNKLITGQESSLGDSPPPEKSE 293
300 PESPMVDNSKNSCODSEADSETSPGFBQEDGSSQSTANKPSRQARDADIEFRKRYST 359
294 PESPMVDNSRNSCODSEADSETSPVFEQDD--RSSQTANKLSSCOAREADGLRKRILT 352
360 KGEVRLHFQFEGGSRRTGMDLNKLPNTSSLNVECRNSKQHGKQKSKITDHLMLPK 419
353 KGEVRLHFQFEGGSRRTGMDLNKLPNTSSLNVECRNSKQHGKQKSKITDHLMLPK 411
420 AEDREKEQWETKHQTERKIPKYVPPHLSPPDKMLGTPIEENMRMPCGIRLPLRPSAN 479
412 SEDRKEQCEVHQHQRTERKIPKYIPNLPPEKKWLTPIEENMRMPCGIRLPLRPSAS 471
480 HVTIVRDLRAGEVPKPPFTHYKDLWNKHVMPCESEONLYPVEDENGERTAGSRWELI 539
472 HVTIVRDLRAGEVPKPPFTHYKDLWNKHVMPCESEONLYPVEDENGERTAGSRWELI 531
540 QTALLNKFTRPQNLKDAILKYNVASKKWDFTALDFWPKVLEAEAAQHLQSLILPDMVK 599
532 QTALLNKFTRPQNLKDAILKYNVASKKWDFTALDFWPKVLEAEAAQHLQSLILPDMVK 591
600 IALCLPNICTQPIPLLKQNNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 659
592 IALCLPNICTQPIPLLKQNNHSITMSQEQIASLLANAFCTFPRRNAMKSEYSSYPDI 651
660 NFNRLEFEGSRSSRKPEKLTLCYFRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTRLH 719
652 NFNRLEFEGSRSSRKPEKLTLCYFRVTEKKPTGLVTFTRQSLDEDPPEWERCEKPLTRLH 711
720 VTYEGTIENGQGMLOVDPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNE 779
712 VTYEGTIENGQGMLOVDPANFVGGVTSAGLVQEEIRFLINPELIIISRLFTEVLHDNE 771
780 CLIIITEQYSEYTGVAETIRWSRSHEDGSDRDCERRCTEIVADILHFRYLDQFVPE 839
772 CLIIITEQYSEYTGVAETIRWSRSHEDGSDRDCERRCTEIVADILHFRYLDQFVPE 831
840 KVRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDARLKALIQIIIAAAAAERDVY 899
832 KVRRELKAYCGFLRPGVSENLSAVATGNWCGAFGGDARLKALIQIIIAAAAAERDVY 891
900 FTFGDSELMRDIYSMHIFLTERKLTGVGVVYKLLRYNEECRNCSTPGPDILKLYPIYHA 959
892 FTFGDSELMRDIYSMHIFLTERKLTGVGVVYKLLRYNEECRNCSTPGPDILKLYPIYHA 951
960 VESCAETADHSGORTGT 976
952 VESCAETADHSGORTGT 976

```

Db 952 VESSAETDMPGQKAGT 968

RESULT 9

US-09-511-507-6

; Sequence 6, Application US/09511507

; Patent No. 6395543

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AME, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/511,507

; CURRENT FILING DATE: 2000-02-23

; PRIOR APPLICATION NUMBER: 09/302,812

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 6

; LENGTH: 968

; TYPE: PRT

; ORGANISM: Mus musculus

; FEATURE:

US-09-511-507-6

Query Match 84.7%; Score 4395; DB 4; Length 968;

Best Local Similarity 85.7%; Pred. No. 0;

Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;

QY 1 MNAAGCEPCTKATWGAA--TTSPAASDARSFSPQRQRLVDPKDAHVQFRVPSPSPACVP 59

Db 1 MSAGPGNEPCTKA-RWGAAGTSAPTADSRSPFQQRRLVDPKDAHVQFRVPSPSPACVS 59

QY 60 GQAGQHRGATSIVFKQKTTISMMDTKGIKTAESISLDSKNNNTNIESMMSSVQKDNFY 119

Db 60 GRAGPHGNATSIVFKQKTTITWMDTKGPKTAESE---SKENNNTIDSMSSVQKDNFY 116

QY 120 QHNVEKLVNUSQLDKSLTEKSTOYLNOHQTAAMCKWQEGKTECOLLESEPTVTLVP 179

Db 117 PHKVEKLVNQLNLDKSPTEKSSQYLNQQTASVCKWQEGKHAEOQLLASEPPAGTLP 176

QY 180 EQFSNANIDRSPQNDHSDTDSBNRDNQOPLTTVKLANAKQTTEDSHAREAKSHQKCSK 239

Db 177 KQLSNANIGSPHTDSDTDEEDRDNQOPLTIKLANKTYPTVGDQAR---SNCKCSG 233

QY 240 SCHFGEDCASCQDEIDVVPKSPISDVGSDDVGTGSKNDNKLIRQESCLGNSPPFEKESE 299

Db 234 SRQSVKDCGQCEVDVLPESPLSDVGAEDIGTGPKNDNKLITQCESLGSPPFEKESE 293

QY 300 PESPMVDNKNKSCQDEADSETSPGDEQEDGSSQOTANKPSRFOARDADIERPKRYST 359

Db 294 PESPMVDNKNKSCQDEADSETSPVDEQDD-RSSQOTANKLSSQARHAGDDURKRYLT 352

QY 360 KGEVRVLFHFOEGESRTGMNDLNAKLPGNITSSLNVECRNSKHGKDKSKITDHLMLPK 419

Db 353 KGEVRVLFHFOE-GENNAGTSDLNKSGNSSSLNVECRNSKHGKDKSKITDHFMAISK 411

QY 420 AEDRRKEQWETKHQTERKIPKYPVPHLSPPDKMLGTPIEMRWMPRCGIHLPLRPSAN 479

Db 412 SEDRRKEQCEVRHQTERKIPKYPVPHLSPPDKMLGTPIEMRWMPRCGIHLPLRPSAS 471

QY 480 HTVTVRVLLRAGEVPKFPPTHYKDLWKNKHVMPKPCSEONLYPVEDENGERTAGSRWLI 539

Db 472 HTVTVRVLLRAGEVPKFPPTHYKDLWKNKHVMPKPCSEONLYPVEDENGERTAGSRWLI 531

QY 540 QTALLNKETRPQNLKADILKKNVAYSKKWDFTALIDFWDKVLEAEAAQHLYSILPDMVK 599

Db 532 QTALLNKETRPQNLKADILKKNVAYSKKWDFTALIDFWDKVLEAEAAQHLYSILPDMVK 591

QY 600 IALCLPNICTQPIPLLKQNMHSITMSQOIASILANAFCTFPRNNAKMKSEYSSYPDI 659

Db 592 IALCLPNICTQPIPLLKQNMHSITMSQOIASILANAFCTFPRNNAKMKSEYSSYPDI 651

QY 660 NFNRLFEGRSSRKPCKLTLFCYFRVRVTEKPTGLVTFTFRQSLQEDPPEWERCCKPLTRLH 719

Db 652 NFNRLFEGRSSRKPCKLTLFCYFRVRVTEKPTGLVTFTFRQSLQEDPPEWERCCKPLTRLH 711

QY 720 VTYEGTIENGOGMLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNE 779

Db 712 VTYEGTIENGOGMLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTFVLDHNE 771

QY 780 CLIIITGTEOYSYVTAETVRSRSHEDSGSDDCERRCTEIVADALHFRRLVDQFVE 839

Db 772 CLIIITGTEOYSYVTAETVRSRSHEDSGSDDCERRCTEIVADALHFRRLVDQFVE 831

QY 840 KMRRELKAYCGFLRPGVSSSENLSAVATCNWCGAGFGGDARLKALIQIIIAAAAEEDVY 899

Db 832 KVRRELKAYCGFLRPGVSENLAVATCNWCGAGFGGDARLKALIQIIIAAAAEEDVY 891

QY 900 FTFGDSELMRDYISMHIFLTERKLTVDYVYKLLRYNNEECNCSCTGPDILKLYPIYHA 959

Db 892 FTFGDSELMRDYISMHIFLTERKLTVDYVYKLLRYNNEECNCSCTGPDILKLYPIYHA 951

QY 960 VESCAETADHSQORTGT 976

Db 952 VESSAETDMPGQKAGT 968

RESULT 10

US-09-302-812-8

; Sequence 8, Application US/09302812B

; Patent No. 6333148

; GENERAL INFORMATION:

; APPLICANT: JACOBSON, Myron K.

; APPLICANT: JACOBSON, Elaine L.

; APPLICANT: AM, Jean-Christophe

; APPLICANT: LIN, Winston

; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ

; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE

; TITLE OF INVENTION: THEREWITH

; FILE REFERENCE: NIAD 201

; CURRENT APPLICATION NUMBER: US/09/302,812B

; CURRENT FILING DATE: 1999-04-30

; EARLIER APPLICATION NUMBER: 60/083,768

; EARLIER FILING DATE: 1998-05-01

; NUMBER OF SEQ ID NOS: 38

; SEQ ID NO 8

; LENGTH: 768

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

; FEATURE:

US-09-302-812-8

Query Match 20.8%; Score 1078; DB 4; Length 768;

Best Local Similarity 42.3%; Pred. No. 7.9e-93;

Matches 240; Conservative 94; Mismatches 184; Indels 50; Gaps 14;

QY 416 RLPKAEADRRKEQWETKHQTERKIPKYPVPHL--SPDKMLGTPIEM--RRMPCGIRL 471

Db 45 RMSKSPDGGISSETEEBE-----PENLANSLSDSWRGVSMEAHHRNQPFLENL 94

QY 472 PLLRPSANHTVTVRVLLRAGEVPKFPPTHYKDLWKNKHVMPKPCSEONLYPVEDENGERT 531

Db 95 PPVTAGNLHVRVYQPLPIRET--PPRYKSPCK--WDSEHVLPCAPESKYPRENPDSTT 150

QY 532 AGSRWELIQTALLNKETRPQNLKADILKKNVAYSKKWDFTALIDFWDKVLEAEAAQHLYS 591

Db 151 IDFRWELIERALLQPIKTCCELQAAIISYNTYEDQWHFRAHQLDLBELDESETRVFFE 210

QY 592 SILPDMVKIALCLPNICTQPIPLLKQNMHSITMSQOIASILANAFCTFPRNNAKMK 650

Db 211 DLLPRIIRLRLPDLIQSPVPLLKHKHNASLSLSSQOISCLLANAFCTFPRNTLKRK 270



[illegible]

RESULT 11  
 US-09-511-477-8  
 ; Sequence 8, Application US/09511477  
 ; Patent NO. 6337202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AME, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIA2 201  
 ; CURRENT APPLICATION NUMBER: US/09/511,477  
 ; CURRENT FILING DATE: 2000-02-23  
 ; PRIOR APPLICATION NUMBER: 09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 ; US-09-511-477-8

```

271 DB SYSTFTPDINFRLYQSTGPAVLEKLCIMHYFRVVCPTERDASNVPTGVVTFVRSGLP 330
703 QY EDFPEWERCEKPL--TRLHVITYEGTIBENGQGLQVDFANRFVGGVTSAGLVQEEIRFL 760
331 DB EHLIDWSQSAAPLGDVPLHVADEGTIEGIGLLQVDFANKYLGGGVLGHGCVQBEIRFV 390
761 QY INPELLISRLFTVELDHNECLIITGEQVSEYTGVAETYRWSRSHEDGSDERDDCERRCTE 820
391 DB ICPELLVVGKLFTECLRPFEALVMLGAERYSNVTGYAGSPWNSGNFEDSTPRDSSGRQTA 450
821 QY IIVADALHFPFRYLDQFVPEKMRRELNKAYCGFLR-----PGVSSSENLSAVATGNWCGGA 874
451 DB IIVADALHFAQSHQVYREDLMERELNKAYIGFVHMVTPPPG-----VATGNWCGGA 502
875 QY FGGDARLKALIQILAAAAAERDVVFTFGDSELMDRIYSMHIFLTERKLTVGDVYKLLIR 934
503 DB FGGDYSYLKALLQIMVCAQIGRPLAYITFGVNEFRDDFHEMALLFRNDGTTVQQLWS-IILR 561
935 QY YNNEECRNCSTGP-----DIKLYPFI 956
562 DB SYSLRIKESKSEPRENKASKKLYDFI 589

RESULT 12
US-09-511-507-8
; Sequence 8, Application US/09511507
; Patent No. 6395543
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AME, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYDROLASE (PARC) EN
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NIAD 201
; CURRENT APPLICATION NUMBER: US/09/511,507
; CURRENT FILING DATE: 2000-02-23
; PRIOR FILING DATE: 09/302,812
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 8
; LENGTH: 768
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; US-09-511-507-8

```

Query Match	20.8%;	Score 1078;	DB 4;	Length 768;
Best Local Similarity	42.3%;	Pred. No. 7.9e-93;		
Matches 240; Conservative	94;	Mismatches 184;	Indels	50; Gaps 14;
QY	416	RLPKAEERKEOWETKHQRTERTKPKVPPHL--SPDKMLGTPIEEM--RRMRPCGIRL	471	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	
	:	: :::	:::	



Db 331 EHLIDWSQSAAPLDGVDPLHDAECTIDEGIGLLQVDFAKYLGGVYLGHCVCQVEIRFV 390  
QY 761 INPELIISRLTFEVLDRNECLIIITGQYSEYTYAETIYRSRSHEDGSDRDCERRCTE 820  
Db 391 ICELLVGLKFTFECRLRFEALVLMGAEKYSNYTGYAGSFWSGNFSTPDSSGRQTA 450  
QY 821 IVAIDALHFRYRLDQFVPEKMRRELKAYCGFLR-----PGVSSSENLSAVATGWCQGA 874  
Db 451 IVAIDALHFAQSHQYREDLMERELKAYIGFVHMVTPPPG-----VATGWCQGA 502  
QY 875 FGDARLAKALIQLIAAARDDVYFEGDSELMRDIYSWHIFLTERKLTGVDVYKLLR 934  
Db 503 FGDYSYLKALLQLMCAQLGRPLAYTYFGNVEFRDDFHEMLLFRNDGTTVQQLWS-ILR 561  
QY 935 YNEECRCNCSPPG-----DIKLYPFI 956  
Db 562 SYSLRIKESKEPRENKASKKLYDFI 589

## RESULT 13

US-09-302-812-10  
; Sequence 10, Application US/09302812B  
; Patent No. 6333148  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) ENZ  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIVE  
; FILE REFERENCE: NIAD 201  
; CURRENT FILING DATE: 1999-04-30  
; EARLIER FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-302-812-10

Query Match 8.9%; Score 463; DB 4; Length 726;  
Best Local Similarity 23.9%; Pred. No. 1.5e-34;  
Matches 207; Conservative 123; Mismatches 278; Indels 258; Gaps 37;

QY 145 YLNQHOTAMCKWQNEGKHITQLLESEPTVTLVPEQFSNANIDRSPQNDHSDTSDSEN 204  
Db 27 FAHQVPTMKRRKLTGHNTHESKED-----PEEPKSRDVFVSSQSDSQSDSAEN 77  
QY 205 -----RDNQOFLT-TVKLANAKO-----TTDEHAREAKSHQKSKCHGEDCASCQ 252  
Db 78 PELAKEVSENCENTLTKISNIESLDNVTERSEHTLD--NH-----KSTPEMEE----- 125  
QY 253 DEIDVVPKSPISDVGSEVDGTGKNDKLIROESCLGNSPPPEKSEPESEPMVDVNSKNS 312  
Db 126 ---DVNKNKI-----DVAINSEDDDELVLLEN-----NKEWR-----DGEQVQQL 163  
QY 313 QDSDRADEE---TSPGFDEQE-----DGSSQTANKPSRFOARDIEF-----RKR 356  
Db 164 SQDLFADQDELTEYFGIMKDTTTLQDLITDSEVETAKQMEMIBETADSTFVGEDSKATNT 223  
QY 357 YSTKG-----GEVRLH-----TFEGGESRTGMNDLNAKLPG----- 388  
Db 224 VRTSSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTG-NLTLQDLINKVDPDRNTRY 282  
QY 389 -----NLSLNVCBRSKQHKKDSKITDHLMLPKAEDRRKEQWETKHQRT-----RK 438  
Db 283 CTIPNPASQGLREDNRGPK-----IVLP-----QRWFEPSRGRDRSYFYFKK 330

QY 439 IPKYVPHLSPDKKWLGTPIEMRRM-PRCGIRLPILRPSANHTVTIRVDLLRAGEVPEK 497  
Db 331 LDGGLKCYKTYGFMFVGLLHNWFEFDPDITYKLPALE----- 368  
QY 498 FPHYKDLWNDKHKVMPKPCSEQLNYPVEDENGERTAGSRWELI--QTALLNKFTRPQNLKDA 556  
Db 369 --MYKEM-----SELVGREVELEKFAVARIA-- 394  
QY 557 ILKYNVAYKKNDFTALIDFWOKVLEAEQAHLVQSLPDMVKIALCLPNICTQPIPLK 616  
Db 395 -----KTAEDILPERIYR-LVGDV----- 412  
QY 617 QKWNHSTMSOEQIASLANAFFCTFPRRNAKMSSESYSPDINFNRLFEGRSSRKPEKL 676  
Db 413 -----ESATLSHKQCAALVARMFF-----ARPSPS-----FCRLSSDKSICVEKL 455  
QY 677 KTLFCYFRRVTEKPTGLVTF--TROSLDF-PWERCERKPLTRLHVT--YEGTIEENGQ 731  
Db 456 KFLFTYFDKMSMDPPDGAVSFRLLTKMDKDTFNEWK--DKLRSLPEVEFFDEMLIEDTA 513  
QY 732 GMLQVDFANRFVGGVTSAGLVQVEIRFLINPELLIISLFTFVLHDNECLITGTEQYSE 791  
Db 514 LCTQVDFANEHLGGVNLHGVSQVEIRFLMCPENMVGMMLCEKWKQLEAISVGAIVPSS 573  
QY 792 YTGVAETVRM-----SRSHEDGSE-RDCCERRCTEIVAIIDALHFR-----YLDQFVPEK 840  
Db 574 YTGCHTLKWAELQPNHSRQNTNEFRDFGLRVELTETDAIDALFKGSKLDCQTEQLNKAN 633  
QY 841 MRRELKAYCGFLRPGVSSSENLSAVATGWCQGAFFGSGARUKALIQLIAAARDDVYF 900  
Db 634 IIREMKKASIGFMSQGPFTNI-DIVTGWGCGAFNGDKPLKFTIQVIAAGVADRLPHFC 692  
QY 901 TFGDSELMRDIYSWHIFLTERKLTVG 926  
Db 693 SFGEPELAACKKIIERMKQKDVTLG 718

## RESULT 14

US-09-511-477-10  
; Sequence 10, Application US/09511477  
; Patent No. 6337202  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT FILING DATE: 1999-04-30  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-477-10

Query Match 8.9%; Score 463; DB 4; Length 726;  
Best Local Similarity 23.9%; Pred. No. 1.5e-34;  
Matches 207; Conservative 123; Mismatches 278; Indels 258; Gaps 37;

QY 145 YLNQHOTAMCKWQNEGKHITQLLESEPTVTLVPEQFSNANIDRSPQNDHSDTSDSEN 204  
Db 27 FAHQVPTMKRRKLTGHNTHESKED-----PEEPKSRDVFVSSQSDSQSDSAEN 77  
QY 205 -----RDNQOFLT-TVKLANAKO-----TTDEHAREAKSHQKSKCHGEDCASCQ 252  
Db 78 PELAKEVSENCENTLTKISNIESLDNVTERSEHTLD--NH-----KSTPEMEE----- 125

253 DEIDVVPKSPVSDVSGSDVGTGSKNDKLRQESCLGNSPPFEKESPEPSPMDVDNSKNS 312  
126 ---DVNKSNI-----DVAINSEDDDELVEEN-----NKEMR-----DGEVQOOL 163  
313 QDSEADEE---TSPGDEQE-----DGSSQTANKPSRFOARDADIEF-----RKR 356  
164 SQDLFADDQELIIEYFGIMKOTTTQDITDSEVETAQKMEMTEETADSTFVGEDSKATKT 223  
357 YSTKG-----GEVRLH-----FQEGGESRTGMNDLNAKLP-----388  
224 VRTSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTEG-NLTLPDLNKVDPDNRVRY 282  
389 ---NISSLNVECRNSKQHKOSKI THLMRLPKAEDRRKEQWETKHORTE-----RK 438  
283 CTIPNFPASQGLREDNRYGPK-----IVLP-----QWRREFDSRRRDSYFYFKRK 455  
439 IPKYVPPHSPDKKWLGPPIEMRRM-PRCGRILPLRPSANHTVIRVDLLRAGEVPK 497  
331 LDGVLKCYKTTGYFMFVGLLHMMEFPDITYKLPAL-----368  
498 PPTHYKDLWKNHVKMPCSEONLYPVEDENGERTAGSRWELT-QTALLNKFRPQNLKDA 556  
369 ---MYKEM-----SELVGREEVLEKPARVARIA--394  
557 ILKYNVAYSKKWDF TALIDFMDKVLLEAQAHLYSILPDMVKIALCLPNICTQPIPLK 616  
395 -----KTAEDILPERIYR-LVGDV-----412  
617 QONHSITMSQOIASILLANAFCTFPRNNAKMSYSSYPDINFNRLPEGRSSRKEPL 676  
413 ---ESATLSHQKCAALVARMF-----ARPDSPFS-----FCRILSSDKSICVEKL 455  
677 KTLFCYFRVTEKPKTGLVTE--TROSLDF-PWERCCEKPLTRLHVT--YEGTIEENQ 731  
456 KFLTFYFDKMSMDPPDGAVSFLTKMDKOTFNEWK--DKLRLSPEVEFFEDMLIEDTA 513  
732 GMLQVDFANRFGVGGVTSAGLQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSE 791  
514 LCTQVDFANEHLGGVNLHSGVQVEIRFLMCEPMVMGMLLCEKMKQLEAISIVGAYVSS 573  
792 YTGAYATYMW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFR-----YLDQFVPEK 840  
574 YTGCHTLKWAELQPNHSRQNTNEFRGRURVETIADAILFKGSKLDCQTEQLNKAN 633  
841 MRRELKAYCGFLRPGVSSNISAVATGNWCGAFGGDARLKALIIQLAAAAARDVVVF 900  
634 IIREMKASIGFMSQGPKEFTNI-PIVTGWMGCGAFNGDKPLKFIIOVIAAGVADRPLHFC 692  
901 TFGDSELMRDIYSMHIFLTERKLTVG 926  
693 SFGEPELAACKCKIIERNKQKOVTLG 718

RESULT 15  
US-09-511-507-10  
; Sequence 10, Application US/09511507  
; Patent No. 6395543  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AME, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOSYLASE (PARG) EN  
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/511,507  
; CURRENT FILING DATE: 2000-02-23  
; PRIOR APPLICATION NUMBER: 09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 10

LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; FEATURE:  
US-09-511-507-10  
Query Match 8.9%; Score 463; DB 4; Length 726;  
Best Local Similarity 23.9%; Pred. No. 1.5e-34;  
Matches 207; Conservative 123; Mismatches 278; Indels 258; Gaps 37;  
QY 145 YLNOHQTAAMCKWQNEKHTTEQLLESFQVTVLTPVEQFSNANIDRSQNDSDHSDTDSSEN 204  
DB 27 FAHQVPMKSRKLTGHTGNTTESK-----PEPKSRDVFVSQSSDESQEDSAEN 77  
QY 205 -----RDNQOFLT-TVLLANAKQ-----TTEDSHAREAKSHQKSKSCHPGEBCASQ 252  
DB 78 PEIAKEVSENCENLTETIKISNIESLONVTERSEHTLD--NH-----KSTEPME-----125  
253 DEIDVVPKSPVSDVSGSDVGTGSKNDKLRQESCLGNSPPFEKESPEPSPMDVDNSKNS 312  
DB 126 ---DVNKSNI-----DVAINSEDDDELVEEN-----NKEMR-----DGEVQOOL 163  
QY 313 QDSEADEE---TSPGDEQE-----DGSSQTANKPSRFOARDADIEF-----RKR 356  
DB 164 SQDLFADDQELIIEYFGIMKOTTTQDITDSEVETAQKMEMTEETADSTFVGEDSKATKT 223  
QY 357 YSTKG-----GEVRLH-----FQEGGESRTGMNDLNAKLP-----388  
DB 224 VRTSSSFLSTVSTCEAPAKGRARMYQKELEKHVIAFTEG-NLTLPDLNKVDPDNRVRY 282  
QY 389 ---NISSLNVECRNSKQHKOSKI THLMRLPKAEDRRKEQWETKHORTE-----RK 438  
DB 283 CTIPNFPASQGLREDNRYGPK-----IVLP-----QWRREFDSRRRDSYFYFKRK 455  
QY 439 IPKYVPPHSPDKKWLGPPIEMRRM-PRCGRILPLRPSANHTVIRVDLLRAGEVPK 497  
DB 331 LDGVLKCYKTTGYFMFVGLLHMMEFPDITYKLPAL-----368  
QY 498 PPTHYKDLWKNHVKMPCSEONLYPVEDENGERTAGSRWELT-QTALLNKFRPQNLKDA 556  
DB 369 ---MYKEM-----SELVGREEVLEKPARVARIA--394  
QY 557 ILKYNVAYSKKWDF TALIDFMDKVLLEAQAHLYSILPDMVKIALCLPNICTQPIPLK 616  
DB 395 -----KTAEDILPERIYR-LVGDV-----412  
QY 617 QONHSITMSQOIASILLANAFCTFPRNNAKMSYSSYPDINFNRLPEGRSSRKEPL 676  
DB 413 ---ESATLSHQKCAALVARMF-----ARPDSPFS-----FCRILSSDKSICVEKL 455  
QY 677 KTLFCYFRVTEKPKTGLVTE--TROSLDF-PWERCCEKPLTRLHVT--YEGTIEENQ 731  
DB 456 KFLTFYFDKMSMDPPDGAVSFLTKMDKOTFNEWK--DKLRLSPEVEFFEDMLIEDTA 513  
QY 732 GMLQVDFANRFGVGGVTSAGLQVEIRFLINPELIIISRLFTVLDHNECLITGTQYSE 791  
DB 514 LCTQVDFANEHLGGVNLHSGVQVEIRFLMCEPMVMGMLLCEKMKQLEAISIVGAYVSS 573  
QY 792 YTGAYATYMW-----SRSHEDGSE-RDDCERRCTEIVAIDALHFR-----YLDQFVPEK 840  
DB 574 YTGCHTLKWAELQPNHSRQNTNEFRGRURVETIADAILFKGSKLDCQTEQLNKAN 633  
QY 841 MRRELKAYCGFLRPGVSSNISAVATGNWCGAFGGDARLKALIIQLAAAAARDVVVF 900  
DB 634 IIREMKASIGFMSQGPKEFTNI-PIVTGWMGCGAFNGDKPLKFIIOVIAAGVADRPLHFC 692  
QY 901 TFGDSELMRDIYSMHIFLTERKLTVG 926  
DB 693 SFGEPELAACKCKIIERNKQKOVTLG 718  
Search completed: May 26, 2004, 18:49:25  
Job time : 36.2745 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2004, 18:46:21 ; Search time 81.4317 Seconds  
(without alignments)  
3343.018 Million cell updates/sec

Title: US-09-302-812-4

Perfect score: 5190  
Sequence: 1 MNAGPCPECTKATRWGAAT.....YHVESCAETADSHGQRTGT 976

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1149313 seqs, 278921704 residues

Total number of hits satisfying chosen parameters: 1149313

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5190	100.0	976	US-09-973-451-4	Sequence 4, Appli
2	4635.5	89.3	977	US-09-973-451-2	Sequence 2, Appli
3	4395	84.7	968	US-09-973-451-6	Sequence 6, Appli
4	1078	20.8	768	US-09-973-451-8	Sequence 8, Appli
5	595	11.5	546	US-10-425-114-60000	Sequence 60000, A
6	463	8.9	726	US-09-973-451-10	Sequence 10, Appli
7	217.5	4.2	200	US-10-424-599-184988	Sequence 184988,
8	188	3.6	180	US-10-424-599-156445	Sequence 156445,
9	179.5	3.5	300	US-10-424-599-233915	Sequence 233915,
10	173.5	3.3	748	US-09-864-761-43244	Sequence 43244, A
11	167	3.2	31	US-09-973-451-11	Sequence 11, Appli
12	166.5	3.2	1163	US-10-221-278-197	Sequence 197, App
13	166.5	3.2	1163	US-10-291-172-197	Sequence 197, App
14	165.5	3.2	1274	US-10-282-122A-43837	Sequence 43837, A
15	164.5	3.2	1111	US-09-815-242-12955	Sequence 12955, A

16	161	3.1	2476	10	US-09-824-574-7	Sequence 7, Appli
17	159	3.1	1462	14	US-10-287-218-17	Sequence 17, Appli
18	158.5	3.1	1948	14	US-10-032-585-7611	Sequence 7611, Ap
19	156.5	3.0	1297	12	US-10-282-122A-71692	Sequence 71692, A
20	155.5	3.0	1163	10	US-09-932-257A-3	Sequence 3, Appli
21	155.5	3.0	1974	15	US-10-369-493-6395	Sequence 6395, Ap
22	154.5	3.0	498	12	US-10-424-599-196154	Sequence 196154,
23	153	2.9	6642	15	US-10-369-493-5013	Sequence 5013, Ap
24	150.5	2.9	408	12	US-10-424-599-224071	Sequence 224071,
25	150.5	2.9	1169	12	US-10-282-122A-71179	Sequence 71179, A
26	150.5	2.9	1312	15	US-10-369-493-1950	Sequence 1950, Ap
27	150.5	2.9	1957	15	US-10-369-493-2070	Sequence 2070, Ap
28	150	2.9	407	12	US-10-282-122A-71177	Sequence 71177, A
29	150	2.9	1501	9	US-09-924-154-17	Sequence 17, Appli
30	149.5	2.9	691	15	US-10-108-260A-4345	Sequence 4345, Ap
31	149.5	2.9	1210	14	US-10-205-823-264	Sequence 264, App
32	149	2.9	26	9	US-09-973-451-20	Sequence 12, Appli
33	149	2.9	29	9	US-09-973-451-12	Sequence 107, App
34	148.5	2.9	665	10	US-09-820-843A-107	Sequence 4552, Ap
35	148.5	2.9	677	14	US-10-060-036-4552	Sequence 1306, Ap
36	148.5	2.9	677	15	US-10-295-027-1306	Sequence 22025, A
37	148.5	2.9	1928	15	US-10-369-493-22025	Sequence 180, App
38	147.5	2.8	677	14	US-10-060-036-180	Sequence 52328, A
39	146.5	2.8	903	12	US-10-282-122A-52328	Sequence 72939, A
40	146.5	2.8	1109	12	US-10-425-114-72939	Sequence 29, Appli
41	146.5	2.8	2125	9	US-09-919-172-29	Sequence 169, App
42	146.5	2.8	2649	14	US-10-205-219-169	Sequence 220, App
43	146.5	2.8	2649	15	US-10-341-434-220	Sequence 230, App
44	146.5	2.8	2649	15	US-10-341-434-230	Sequence 15, Appli
45	145.5	2.8	1908	16	US-10-197-824-15	

#### ALIGNMENTS

##### RESULT 1

```
US-09-973-451-4
; Sequence 4, Application US/09973451
; Patent No. US20020132328A1
; GENERAL INFORMATION:
; APPLICANT: JACOBSON, Myron K.
; APPLICANT: JACOBSON, Elaine L.
; APPLICANT: AM, Jean-Christophe
; APPLICANT: LIN, Winston
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE
; TITLE OF INVENTION: (PARG) ENZYMES
; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: NTAD 201
; CURRENT APPLICATION NUMBER: US/09/973,451
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: US/09/302,812
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 60/083,768
; PRIOR FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 4
; LENGTH: 976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
US-09-973-451-4
```

Query Match 100.0% Score 5190; DB 9; Length 976;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 976; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNAGPCPECTKATRWGAATTS PAASDARSFPSPRRVLDPKDAHVQFRVPSSPACVPG 60

Db 1 MNAGPCPECTKATRWGAATTS PAASDARSFPSPRRVLDPKDAHVQFRVPSSPACVPG 60

QY 61 QAGQHRGSAISLVFKQKTTSMWDTKGTAESELSKNNNTIESMWSVQKDNFYQ 120

Thu May 27 09:56:05 2004

Db 61 QAGHRSATSLVFKQKTIISWMDTKGKTAESBLSKNNNTRIESMSSVQKDFYQ 120  
Qy 121 HNEKLVNVSQSLDKSLSTKSTQVYNQHOQTAAMCKQWNEKGHTTEOLLESEPTVTLVPE 180  
Db 121 HNEKLVNVSQSLDKSLSTKSTQVYNQHOQTAAMCKQWNEKGHTTEOLLESEPTVTLVPE 180  
Qy 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTIVKLANAKOTTIDEHAREAKSHQKCSKS 240  
Db 181 QFSNANIDRSQNDHSDTDSEENRDNQOFLTIVKLANAKOTTIDEHAREAKSHQKCSKS 240  
Qy 241 CHPGEDCASCQODEIDVVPKSPSLSDVSGEDVGTGSKNDKLIROBSCIGNSPPEKESEP 300  
Db 241 CHPGEDCASCQODEIDVVPKSPSLSDVSGEDVGTGSKNDKLIROBSCIGNSPPEKESEP 300  
Qy 301 ESPMDVDNKSQCQSEADEETSPGDEQDSSQOTANKPSRFOARDADIEFRKRYSTK 360  
Db 301 ESPMDVDNKSQCQSEADEETSPGDEQDSSQOTANKPSRFOARDADIEFRKRYSTK 360  
Qy 361 GGEVRLHFOFEGGESRTGMNDLNKLPNISSINVECRNSKQHGKDKSITDHLMLPKA 420  
Db 361 GGEVRLHFOFEGGESRTGMNDLNKLPNISSINVECRNSKQHGKDKSITDHLMLPKA 420  
Qy 421 EDRRKEQWETKHQRTKIPKTVPPHLSPLDKKWLGTPIEEMRMPCRGIRLPLRPSANH 480  
Db 421 EDRRKEQWETKHQRTKIPKTVPPHLSPLDKKWLGTPIEEMRMPCRGIRLPLRPSANH 480  
Qy 481 TVTIRVDLLRAGEVPEKPTTHYKDLWKNKHVMPCEQNLVPEDENGERTAGSRWELIQ 540  
Db 481 TVTIRVDLLRAGEVPEKPTTHYKDLWKNKHVMPCEQNLVPEDENGERTAGSRWELIQ 540  
Qy 541 TALLNKFTPNQDKAILKYNVAYSKKWDFLTALIDFWDKVLBEAEAAHLYQSILPDMVKI 600  
Db 541 TALLNKFTPNQDKAILKYNVAYSKKWDFLTALIDFWDKVLBEAEAAHLYQSILPDMVKI 600  
Qy 601 ALCLPNICTQPIPLKQKMNHSITMSQEQIASILANAFCTPPRRNAKMKSEYSSYPDIN 660  
Db 601 ALCLPNICTQPIPLKQKMNHSITMSQEQIASILANAFCTPPRRNAKMKSEYSSYPDIN 660  
Qy 661 FNLFPGRSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDDFPWEKCEKPLTRLHV 720  
Db 661 FNLFPGRSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDDFPWEKCEKPLTRLHV 720  
Qy 721 TYEGTTEENQGMQLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLHNEC 780  
Db 721 TYEGTTEENQGMQLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLHNEC 780  
Qy 781 LIITGTEQYSEYTGIAETYNWSHEDGSRDCCERRCTEIVDAIDALHFRYLDQFVPEK 840  
Db 781 LIITGTEQYSEYTGIAETYNWSHEDGSRDCCERRCTEIVDAIDALHFRYLDQFVPEK 840  
Qy 841 MRBELNKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 900  
Db 841 MRBELNKAYCGFLRPGVSSNLSAVATGNWCGGAFGGDARLKALIQILAAAAAERDVVYF 900  
Qy 901 TFGDSELMRDLYSMHIFLTERKLTGVGVYKLLRYNNECRNSTPGPDIKLPFIYHAV 960  
Db 901 TFGDSELMRDLYSMHIFLTERKLTGVGVYKLLRYNNECRNSTPGPDIKLPFIYHAV 960  
Qy 961 ESCAETADHSGQRTGT 976  
Db 961 ESCAETADHSGQRTGT 976

RESULT 2  
US-09-973-451-2  
; Sequence 2, Application US/09973451  
; Patent No. US20020132328A1  
; GENERAL INFORMATION:  
; APPLICANT: JACOBSON, Myron K.  
; APPLICANT: JACOBSON, Elaine L.  
; APPLICANT: AM, Jean-Christophe  
; APPLICANT: LIN, Winston  
; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE

; TITLE OF INVENTION: (PARG) ENZYMES, THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
; TITLE OF INVENTION: THEREWITH  
; FILE REFERENCE: NIAD 201  
; CURRENT APPLICATION NUMBER: US/09/973,451  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US/09/302,812  
; PRIOR FILING DATE: 1999-04-30  
; PRIOR APPLICATION NUMBER: 60/083,768  
; PRIOR FILING DATE: 1998-05-01  
; NUMBER OF SEQ ID NOS: 38  
; SEQ ID NO 2  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Bos taurus  
; FEATURE:  
US-09-973-451-2

Query Match 89.3%; Score 4635.5; DB 9; Length 977;  
Best Local Similarity 89.5%; Pred. No. 0;  
Matches 874; Conservative 32; Mismatches 69; Indels 1; Gaps 1;  
Qy 1 MNAGPCPECTKATRWGAATTS-PAASDARSPFRQREVLPKDAHVQFRVPPSPACVP 59  
Db 1 MSAGPCPECTKRPWDAATTSPPAASDARSPFRQREVLPKDAHVQFRVPPSPSGCAL 60  
Qy 60 GOAGHRSATSLVFKQKTIISWMDTKGKTAESBLSKNNNTRIESMSSVQKDFY 119  
Db 61 GRAGHRSATSLVFKQKTIISWMDTKGKTVESBLSKNNNTRIESMSSVQKDFY 120  
Qy 120 QHNVKLVNVSQSLDKSLSTKSTQVYNQHOQTAAMCKQWNEKGHTTEOLLESEPTVTLV 179  
Db 121 QENMEKLVNVSQSLGFDKSPVEKGTQYLKQHOQTAAMCKQWNEGHPHERLLESEPPAVTLV 180  
Qy 180 EGFNSNANIDRSQNDHSDTDSEENRDNQOFLTIVKLANAKOTTIDEHAREAKSHQKCSK 239  
Db 181 EGFNSNANIDRSQNDHSDTDSEENRDNQOFLTIVKLANAKOTTIDEHAREAKSHQKCSK 240  
Qy 240 SCHPGEDCASCQODEIDVVPKSPSLSDVSGEDVGTGSKNDKLIROBSCIGNSPPEKESE 299  
Db 241 ACHPAEACAGCQOEDTVVSESLSDTSGEDVGTGLKNNANLNQDESSIGNSPPEKESE 300  
Qy 300 PESPMDVDNKSQCQSEADEETSPGDEQDSSQOTANKPSRFOARDADIEFRKRYST 359  
Db 301 PESPMDVDNKSQCQSEADEETSPGDEQDSSQOTANKPSRFOARDADIEFRKRYST 360  
Qy 360 KGEVRLHFOFEGGESRTGMNDLNKLPNISSINVECRNSKQHGKDKSITDHLMLPK 419  
Db 361 KGEVRLHFOFEGGESRTGMNDLNKLPNISSINVECRNSKQHGKDKSITDHLMLPK 420  
Qy 420 AEDRRKEQWETKHQRTKIPKTVPPHLSPLDKKWLGTPIEEMRMPCRGIRLPLRPSAN 479  
Db 421 AEDRRKEQWETKHQRTKIPKTVPPHLSPLDKKWLGTPIEEMRMPCRGIRLPLRPSAN 480  
Qy 480 HTVTIRVDLLRAGEVPEKPTTHYKDLWKNKHVMPCEQNLVPEDENGERTAGSRWELI 539  
Db 481 HTVTIRVDLLRAGEVPEKPTTHYKDLWKNKHVMPCEQNLVPEDENGERTAGSRWELI 540  
Qy 540 QTALLNKFTPNQDKAILKYNVAYSKKWDFLTALIDFWDKVLBEAEAAHLYQSILPDMVK 599  
Db 541 QTALLNKFTPNQDKAILKYNVAYSKKWDFLTALIDFWDKVLBEAEAAHLYQSILPDMVK 600  
Qy 600 IALCLPNICTQPIPLKQKMNHSITMSQEQIASILANAFCTPPRRNAKMKSEYSSYPDI 659  
Db 601 IALCLPNICTQPIPLKQKMNHSITMSQEQIASILANAFCTPPRRNAKMKSEYSSYPDI 660  
Qy 660 NENRLEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDDFPWEKCEKPLTRLH 719  
Db 661 NENRLEGRSSRKPEKLTLCYFRRVTEKPTGLVTFTRQSLDDFPWEKCEKPLTRLH 720  
Qy 720 VTYEGTTEENQGMQLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLHNE 779  
Db 721 VTYEGTTEENQGMQLQVDFANRFGVGGVTSAGLVQEEIRFLINPELIIISRLFTVLHNE 780

us-09-302-812-4.rapb

Thu May 27 09:56:05 2004

QY 780 CLITGTQYSEYTGAEYTRWSRSHEDSBERDCCERCTEIVADALHFRYLDQFVPE 839  
 DB 781 CLITGTQYSEYTGAEYTRWSRSHEDSBERDCCERCTEIVADALHFRYLDQFVPE 840  
 QY 840 KMRRLNKAYCGFLRPGVSSNLAVATGNWCGAGFGDARLKALITQILAAAAAERDVVY 899  
 DB 841 KMRRLNKAYCGFLRPGVSSNLAVATGNWCGAGFGDARLKALITQILAAAAAERDVVY 900  
 QY 900 FTFGDSELMRDYISMHIFITERKLTVDGVYKLLRYNEECRNCSTPGDPIKLYPIIYHA 959  
 DB 901 FTFGDSELMRDYISMHIFITERKLTVDGVYKLLRYNEECRNCSTPGDPIKLYPIIYHA 960  
 QY 960 VESCAETADHSGORTG 975  
 DB 961 VESCTQTQNGORTG 976  
 RESULT 3  
 US-09-973-451-6  
 ; Sequence 6, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; PRIOR FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 6  
 ; LENGTH: 968  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; FEATURE:  
 US-09-973-451-6  
 Query Match 84.7%; Score 4395; DB 9; Length 968;  
 Best Local Similarity 85.7%; Pred. No. 0;  
 Matches 837; Conservative 52; Mismatches 78; Indels 10; Gaps 6;  
 QY 1 MNAGPGCEPCTKATRWGAA-TTSPAASDARSPRRORRVLDPKDAHVQFRVPPSPACVP 59  
 DB 1 MSAGPGCEPCTKATRWGAA-TTSPAASDARSPRRORRVLDPKDAHVQFRVPPSPACVP 59  
 QY 60 GOAGHGRSATSPLVPKQITTSWMDTKGKTAESESLDSKENNTRIDSMSSVQKONFY 119  
 DB 60 GRAGHGRNATSPLVPKQITTSWMDTKGKTAESESLDSKENNTRIDSMSSVQKONFY 116  
 QY 120 QHNVEKLVNSQLSDKSLTEKSTOYLQHQTAAMCKQWNEKHTEQLLESPQVTLVP 179  
 DB 117 PHKVEKLVNPQLNDKSPTEKSSQVNLQOQTASVCKWQNEKHAEQLLASEPPAGTFLP 176  
 QY 180 EGFNSANIDRSQNDHSDTDEENRDNQOFLTIVKLANAKOTTEDEHAREAKSHQKCSK 239  
 DB 177 KQLSNANIGQSPTDHSVTDEEDRDNQOFLTIVKLANAKOTTEDEHAREAKSHQKCSK 233  
 QY 240 SCHPGEDCASCQDEIDVVPKPLSDVSGEDVGTGSKNDKLIROESCLGNSPPPEKSE 299  
 DB 234 SQQSVKDTGCCQEEVDVLPESPLSDVGAEDIGTGFKNDKLTGQESSLSDSPPEKSE 293  
 QY 300 PESPMVDNRSKNSQDSEADETSPGDEQDGSSQTKANKPSRFQARADIEFRKRYST 359  
 DB 294 PESPMVDNRSKNSQDSEADETSPGDEQDGSSQTKANKPSRFQARADIEFRKRYST 352

QY 360 KGGEVRLHFQEGGSRRTGMMDLNALKPLGNISSINVECRNSKQHKOSKITDHLMLPK 419  
 DB 353 KGGEVRLHFQEGGSRRTGMMDLNALKPLGNISSINVECRNSKQHKOSKITDHLMLPK 411  
 QY 420 AEDRKEQWETHQORTERKIPKYVPPHLSPDCKWGLGTPLEENRMRPRCGIRLFLRPSAN 479  
 DB 412 SEDRRKEQCEVHQORTERKIPKYVPPHLSPDCKWGLGTPLEENRMRPRCGIRLFLRPSAN 471  
 QY 480 HTVTTRVOLLRAGEVPKPPPTHYKDLMDNKHVMPSCSEONLYPVEDENGERTAGSRWELI 539  
 DB 472 HTVTTRVOLLRAGEVPKPPPTHYKDLMDNKHVMPSCSEONLYPVEDENGERTAGSRWELI 531  
 QY 540 QTALLNKFTPRONLKDAILKYNVAYSKKWDFTALVDKVIIEEAQAQHLYSILPDVVK 599  
 DB 532 QTALLNKFTPRONLKDAILKYNVAYSKKWDFTALVDKVIIEEAQAQHLYSILPDVVK 591  
 QY 600 IALCLPNICTQPIPLKQKQNHSTWMSQOIASLLANAFCTPPRNNAKMKSEYSPDI 659  
 DB 592 IALCLPNICTQPIPLKQKQNHSTWMSQOIASLLANAFCTPPRNNAKMKSEYSPDI 651  
 QY 660 NFNRLFEGSSRRKPEKLTFCYFRVTEKPTGLVTFTQSLDPEFWECEKPLTRLH 719  
 DB 652 NFNRLFEGSSRRKPEKLTFCYFRVTEKPTGLVTFTQSLDPEFWECEKPLTRLH 711  
 QY 720 VTYEGTIENGQMLQVDFANRFVGGVTSAGLVOEIEIRFLINPELISRLFTVLDHNE 779  
 DB 712 VTYEGTIENGQMLQVDFANRFVGGVTSAGLVOEIEIRFLINPELISRLFTVLDHNE 771  
 QY 780 CLITGTQYSEYTGAEYTRWSRSHEDSBERDCCERCTEIVADALHFRYLDQFVPE 839  
 DB 772 CLITGTQYSEYTGAEYTRWSRSHEDSBERDCCERCTEIVADALHFRYLDQFVPE 831  
 QY 840 KMRRLNKAYCGFLRPGVSSNLAVATGNWCGAGFGDARLKALITQILAAAAAERDVVY 899  
 DB 832 KMRRLNKAYCGFLRPGVSSNLAVATGNWCGAGFGDARLKALITQILAAAAAERDVVY 891  
 QY 900 FTFGDSELMRDYISMHIFITERKLTVDGVYKLLRYNEECRNCSTPGDPIKLYPIIYHA 959  
 DB 892 FTFGDSELMRDYISMHIFITERKLTVDGVYKLLRYNEECRNCSTPGDPIKLYPIIYHA 951  
 QY 960 VESCAETADHSGORTG 976  
 DB 952 VESCAETADHSGORTG 968  
 RESULT 4  
 US-09-973-451-8  
 ; Sequence 8, Application US/09973451  
 ; Patent No. US20020132328A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACOBSON, Myron K.  
 ; APPLICANT: JACOBSON, Elaine L.  
 ; APPLICANT: AM, Jean-Christophe  
 ; APPLICANT: LIN, Winston  
 ; TITLE OF INVENTION: GENES ENCODING SEVERAL POLY (ADP-RIBOSE) GLYCOHYDROLASE  
 ; TITLE OF INVENTION: (PARG) ENZYMES,  
 ; TITLE OF INVENTION: THE PROTEINS AND FRAGMENTS THEREOF, AND ANTIBODIES IMMUNOREACTIV  
 ; TITLE OF INVENTION: THEREWITH  
 ; FILE REFERENCE: NIAD 201  
 ; CURRENT APPLICATION NUMBER: US/09/973,451  
 ; PRIOR FILING DATE: 2001-10-09  
 ; PRIOR APPLICATION NUMBER: US/09/302,812  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: 60/083,768  
 ; PRIOR FILING DATE: 1998-05-01  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SEQ ID NO 8  
 ; LENGTH: 768  
 ; TYPE: PRT  
 ; ORGANISM: Drosophila melanogaster  
 ; FEATURE:  
 US-09-973-451-8